



Titre: Identification de la relation sémantique sous-jacente des noms
Title: composés

Auteur: Thierry Bélair
Author:

Date: 2014

Type: Mémoire ou thèse / Dissertation or Thesis

Référence: Bélair, T. (2014). Identification de la relation sémantique sous-jacente des noms
Citation: composés [Mémoire de maîtrise, École Polytechnique de Montréal]. PolyPublie.
<https://publications.polymtl.ca/1611/>

 **Document en libre accès dans PolyPublie**
Open Access document in PolyPublie

URL de PolyPublie: <https://publications.polymtl.ca/1611/>
PolyPublie URL:

**Directeurs de
recherche:** Michel Gagnon, & Caroline Barrière
Advisors:

Programme: Génie informatique
Program:

UNIVERSITÉ DE MONTRÉAL

IDENTIFICATION DE LA RELATION SÉMANTIQUE SOUS-JACENTE DES NOMS
COMPOSÉS

THIERRY BÉLAIR
DÉPARTEMENT DE GÉNIE INFORMATIQUE ET GÉNIE LOGICIEL
ÉCOLE POLYTECHNIQUE DE MONTRÉAL

MÉMOIRE PRÉSENTÉ EN VUE DE L'OBTENTION
DU DIPLÔME DE MAÎTRISE ÈS SCIENCES APPLIQUÉES
(GÉNIE INFORMATIQUE)
NOVEMBRE 2014

UNIVERSITÉ DE MONTRÉAL

ÉCOLE POLYTECHNIQUE DE MONTRÉAL

Ce mémoire intitulé :

IDENTIFICATION DE LA RELATION SÉMANTIQUE SOUS-JACENTE DES NOMS
COMPOSÉS

présenté par : BÉLAIR Thierry

en vue de l'obtention du diplôme de : Maîtrise ès sciences appliquées

a été dûment accepté par le jury d'examen constitué de :

M. DESMARAIS Michel C., Ph.D., président

M. GAGNON Michel, Ph.D., membre et directeur de recherche

Mme BARRIÈRE Caroline, Ph.D., membre et codirectrice de recherche

Mme KOSSEIM Leila, Ph.D., membre

DÉDICACE

*Je dédie ce mémoire à deux orchestres de Montréal,
l'Orchestre de jeux vidéos et l'Orchestre à vent de musique de film,
puisque'ils ont rendu mélodieuse sa rédaction.*

REMERCIEMENTS

Merci à mes collègues du Centre de recherche informatique de Montréal pour l'accueil chaleureux.

Ce travail n'aurait pas été possible sans l'apport financier du gouvernement provincial et fédéral et de l'entreprise 3CE (www.3ceonline.com).

Merci à Chistophe Botek, Konstantinos Lambrou-Latreille, Caroline Barrière et Michel Gagnon pour les longues heures d'annotation de noms composés.

Merci aux associations étudiantes qui défendent constamment mes droits d'étudiant universitaire.

Ma gratitude pour Louise Demers dépasse le cadre de ce simple mémoire, mais je tiens à la remercier d'avoir été ma guide et mon modèle dans les études aux cycles supérieurs.

RÉSUMÉ

Les descriptions de produits d’inventaire, dans l’industrie d’importation/exportation, comptent beaucoup de noms composés, des paires de noms ayant une signification particulière. Nous pourrions mieux comprendre la description, et classer le produit, s’il était possible de comprendre les noms composés dans leur ensemble : deux noms et une relation qui les relie, implicite. Par exemple, le nom composé *olive oil* peut être interprété avec la paraphrase *oil that comes from olive*, qui rend explicite la relation de provenance entre *olive* et *oil*. Selon Levi (Levi, 1978), douze relations sémantiques sont possibles pour un nom composé.

Dans cette recherche, nous reproduisons presque en totalité l’expérience de Nakov (Nakov and Hearst, 2008). À partir d’un nom composé, il génère des requêtes à Google pour obtenir des textes où les deux noms du nom composé sont en relation. Les verbes et particules qui relient les deux noms sont regroupés dans une liste d’éléments prédicatifs, appelée vecteur prédicatif. Le nom composé est ensuite comparé à d’autres noms composés dont la relation est connue, les noms composés de référence.

Malheureusement, depuis 2011, la collecte des textes d’internet en utilisant Google n’est plus possible. Notre objectif est donc de rendre indépendant du moteur de recherche de Google l’identification de la relation de nom composé en maintenant le même taux de succès pour l’identification de relation que Nakov. Nous faisons l’hypothèse que nous pouvons identifier la relation sémantique grâce aux textes trouvés sur Wikipédia, sur Faroo, sur Yahoo ou dans les n-grammes de Google.

Nous modifions le protocole en plusieurs points. Nous utilisons d’autres corpus. Nous enrichissons la requête avec différents synonymes. Nous refaisons l’expérience avec d’autres vecteurs de référence. Nous modifions la comparaison entre vecteurs de référence et vecteurs du corpus. Finalement, nous essayons aussi des façons différentes de choisir la relation sémantique à partir des résultats précédents.

Nous testons le succès de notre prototype de trois façons : en utilisant les noms composés de référence (par *Leave-one-out*), en utilisant les mêmes noms composés, mais reclassifiés et en utilisant 417 nouveaux noms composés.

Nous atteignons au mieux le taux de succès de 33% alors que Nakov rapporte celui de 43%.

Malgré les nombreuses possibilités d’amélioration, nos résultats actuels sont significativement en dessous des résultats de la littérature. Il nous faut conclure que dans l’état actuel, l’implémentation de l’approche de Nakov sans le grand nombre de documents indexés par Google

n'est pas suffisante pour la catégorisation de noms composés parmi 12 relations implicites.

ABSTRACT

The import/export business must keep track of many different products, but the descriptions of the products are often disorganized. For a machine to decipher the descriptions better, noun compounds could be used. They are numerous in the descriptions. Noun compounds (NC) are a pair of consecutive nouns that have acquired a new meaning. The two nouns are linked by a semantic relation. For example, the noun compound *olive oil* can be interpreted as *oil that comes from olive*. According to Levi, there are 12 possible relations for noun compounds, including **FROM**.

In this work, we replicated how Nakov attributed one of the 12 relations to noun compounds. Starting with an unknown noun compound, requests are automatically sent to Google in order to retrieve documents in which the two nouns appear. The features that link the two nouns are extracted and collected into a feature vector. A feature can be either a verb or a verb with its accompanying particle. By comparing feature vectors, the noun compound can be compared to others noun compounds whose relations are known.

Unfortunately, since 2011 Google does not accept automatically generated requests. Our goal is to adapt the approach of Nakov to identify the semantic relation in a noun compound without using Google Search. We assert that it is possible by using the texts of Wikipedia, Faroo, Yahoo or the Google n-grams. We modified the approach in a few ways. In addition to using a different text corpus, we form new noun compounds from the synonyms of the first noun compound. The experiment is attempted with feature vectors tweaked in different ways. We try several ways to compare the noun compounds and choose the semantic relation outputted by the system.

Our prototype was submitted to three separate evaluations: one using the known noun compounds with the Leave-one-out approach, a second using the same noun compounds, but with different relations and a third using 417 new noun compounds.

At best, our success ratio reaches 33%, whereas Nakov reports achieving a 43% success ratio.

Even though there are many ways to improve this result, our results are subpar compared to the state of the art. We must conclude that, using a smaller corpus, the method of Nakov cannot correctly classify noun compounds into one of 12 possible semantics relation.

TABLE DES MATIÈRES

DÉDICACE	iii
REMERCIEMENTS	iv
RÉSUMÉ	v
ABSTRACT	vii
TABLE DES MATIÈRES	viii
LISTE DES TABLEAUX	xi
LISTE DES FIGURES	xii
LISTE DES SIGLES ET ABRÉVIATIONS	xiii
LISTE DES ANNEXES	xv
CHAPITRE 1 INTRODUCTION	1
CHAPITRE 2 SÉMANTIQUE DES NOMS COMPOSÉS	4
2.1 Noms composés	4
2.2 Prédicats	7
2.2.1 Prédicats pour paraphraser les noms composés	7
2.2.2 Prédicats proposés par Levi	8
2.2.3 Vecteur prédictif	11
CHAPITRE 3 APPROCHES POUR L'INTERPRÉTATION DES NOMS COMPOSÉS	12
CHAPITRE 4 TECHNIQUE D'EXTRACTION DE LA RELATION DE NAKOV .	16
4.1 Description générale	16
4.2 Extraction du corpus	17
4.3 Analyse	19
4.4 Vecteurs prédictifs de référence	20
4.5 Déduction de la relation	22
4.6 Évaluation	23

CHAPITRE 5	MÉTHODOLOGIE	26
5.1	Description générale	26
5.2	Extraction du corpus	26
5.2.1	Description des corpus utilisés	27
5.2.2	Préparation de phrases en vue de l'analyse	29
5.2.3	Expansion par synonymes	30
5.3	Patrons	32
5.3.1	Éléments prédictifs au final	32
5.3.2	Analyse syntaxique de surface	33
5.3.3	Les patrons et leur expression régulière	34
5.3.4	Différents types de patrons utilisés	37
5.3.5	Regroupements de patrons	40
5.4	Vecteurs prédictifs de référence	42
5.4.1	Normalisation et nettoyage des éléments prédictifs	43
5.4.2	Choix des éléments prédictifs de référence	46
5.5	Identification de la relation	47
5.5.1	Extraction de la relation à partir du voisinage	48
5.5.2	Analyse des vecteurs pour prédire quel voisinage choisir	49
5.6	Résumé des différences avec la méthode de Nakov	51
CHAPITRE 6	RÉSULTATS ET ANALYSE	53
6.1	Points de comparaison et résultat	53
6.1.1	Comparaison avec d'autres approches	54
6.1.2	Avec résultats intermédiaires de Nakov sur Google	54
6.2	Différentes méthodes d'évaluation	55
6.2.1	Leave-one-out	55
6.2.2	Évaluation avec le corpus indépendant de Kim	56
6.2.3	Recatégorisation de Levi	57
6.2.4	Résultats aux différentes méthodes d'évaluation	58
6.3	Résultats pour les variations du protocole	61
6.3.1	Variation du corpus	61
6.3.2	Approche par synonyme	62
6.3.3	Équipe de patrons	65
6.3.4	Variation dans la sélection et le traitement des éléments prédictifs de référence	67
6.3.5	Méthode d'identification de la relation sémantique	69

6.4	Analyses supplémentaires	70
6.4.1	Erreurs dans l'extraction des éléments prédicatifs	70
6.4.2	Influence des vecteurs prédicatifs	71
6.4.3	Influence induite de la relation BE	72
6.5	Conclusions	73
CHAPITRE 7 CONCLUSION ET TRAVAUX FUTURS		77
RÉFÉRENCES		80
ANNEXES		85

LISTE DES TABLEAUX

Tableau 5.1	Séparateurs de phrases pour différents corpus	30
Tableau 5.2	Contributions des types de patrons	42
Tableau 5.3	Meilleurs regroupements de patrons selon l'expérience de sélection . .	42
Tableau 5.4	Normalisation	45
Tableau 5.5	Intersection moyenne entre les différentes relations de Levi	50
Tableau 6.1	Répartition des assignations de prédicats de Levi aux noms composés de Kim sélectionnés	57
Tableau 6.2	Répartition des nouvelles assignations de prédicats pour les noms com- posés de Levi	58
Tableau 6.3	Résultats pour différentes méthodes d'évaluation	59
Tableau 6.4	Statistiques de performance des différents corpus	62
Tableau 6.5	Taux de succès pour différents ensembles de synonymes en requête au corpus	63
Tableau 6.6	Taux de succès pour différents patrons	66
Tableau 6.7	Quantité de documents et de phrases pour les différents corpus avec les patrons de Nakov	67
Tableau 6.8	Taux de succès pour différentes sélections des éléments précatifs . .	68
Tableau 6.9	Taux de succès pour différentes méthodes d'identification de la relation sémantique	70
Tableau 6.10	Métriques pour les vecteurs d'éléments précatifs trouvés pour les 212 noms composés	72
Tableau 6.11	Succès selon les relations sémantiques	73
Tableau 6.12	Mauvaises attributions selon les relations sémantiques	74
Tableau D.1.	Tous les patrons d'extraction d'éléments précatifs utilisés	133

LISTE DES FIGURES

Figure 4.1	Méthode de Nakov de base	17
Figure 5.1	Extraction de contexte à partir du corpus	27
Figure 5.2	Différents ensembles de patrons possibles pour l'extraction d'éléments prédicatifs : <i>trait plein</i> : Nakov, <i>trait pointillé</i> : meilleur rappel	38
Figure 5.3	Positionnement de la normalisation et du nettoyage des éléments pré- dicatifs	44

LISTE DES SIGLES ET ABRÉVIATIONS

3CE	la compagnie 3CE (www.3ceonline.com) (Commodity Code Classification Engine)
B-NP	annotation pour un mot au début d'un syntagme nominal (expression du Penn Treebank Project)
CR	patrons pour extraire les éléments prédicatifs de clauses relatives (Clauses Relatives)
GP	patrons pour extraire les éléments prédicatifs malgré la présence de groupes prépositionnels entre le groupe nominal et le groupe verbal (Groupe Prépositionnel)
HS	Harmonized System (écrit par l'organisation mondiale des douanes)
IN	annotation de catégorie pour une conjonction subordonnée ou une préposition (expression du Penn Treebank Project)
I-NP	annotation pour un mot au milieu ou à la fin d'un syntagme nominal (expression du Penn Treebank Project)
MeSH	documentation pour la médecine utilisé par Rosario et Hearst
N ₁	nom #1, premier mot du nom composé
N ₂	nom #2, deuxième mot du nom composé
NC	nom composé ou <i>noun compound</i>
N-Gram	les n-grammes web de Google, des ensembles de mots accompagnés de leur fréquence d'apparition sur le web
NLP	<i>Natural Language Processing</i> (traduit par Traitement automatique des langues naturelles)
NP	syntagme nominal (<i>noun phrase</i>)
OMD	Organisation Mondiale des Douanes (<i>World Custom Organization</i>)
PART	patrons pour extraire les éléments prédicatifs avec particules
PASS	patrons pour extraire les éléments prédicatifs de phrases passives
RP	annotation de catégorie pour une particule ou un adverbe (expression du Penn Treebank Project)
SemEval	compétition de traitement de langues naturelles
SH	Système harmonisé (écrit par l'OMD, aussi HS)
TALN	Traitement automatique des langues naturelles

TF-IDF	mesure mathématique qui balance la fréquence dans un contexte et la rareté dans un autre contexte, plus général (pour l'anglais Term Frequency * Inverse Document Frequency).
TS	patrons pour extraire les éléments prédicatifs seulement si N_1 et N_2 sont à la tête de leur groupe (Tête seulement)
VB	annotation pour un verbe (<i>Verb</i>) (expression du Penn Treebank Project)
VP	(<i>verbal phrase</i>) (expression du Penn Treebank Project)
WCO	<i>World Custom Organization</i>
WDT	annotation de catégorie pour un déterminant en <i>wh</i> - (expression du Penn Treebank Project)
WP	annotation de catégorie pour un pronom personnel en <i>wh</i> - (expression du Penn Treebank Project)

LISTE DES ANNEXES

Annexe A	Noms composés de Levi	85
Annexe B	Noms composés de Levi recatégorisés	89
Annexe C	Vecteurs prédicatifs de référence de Mturk fournis par Nakov	93
Annexe D	Patrons d'extraction	133
Annexe E	Noms composés de Kim classés selon les 12 prédicats de Levi	138

CHAPITRE 1 INTRODUCTION

Les exportateurs sont directement affectés par les tarifs de douane. Il faut pouvoir déterminer quel montant devra être payé pour faire voyager leurs produits. Pour ce faire, ils utilisent le système harmonisé (SH) qui est défini par l'Organisation mondiale des douanes. Le système est utilisé dans 98% des pays pour permettre l'identification des produits d'exportation sans être affecté par les changements dans les tarifs de douane. Les différents pays décident des tarifs d'importation appliqués à chaque catégorie définie dans le SH. Le SH se base sur la matière première des produits pour trouver une catégorie à toute la variété de produits déplaçables.

Trouver la bonne catégorie du SH pour tous les produits en stock est une tâche difficile pour un exportateur. Elle est rendue plus difficile encore si la description des nombreux produits ressemble à `5/8"x6', 2 "leg wire rope bridle", thimble w/ "aluminum tube tag" (all pull test certified)`¹. Des entreprises comme 3CE sont engagées pour faciliter la classification. Elles appliquent l'intelligence artificielle pour classer des milliers de descriptions d'inventaire souvent précises, mais sans contexte. Elles sont parfois ambiguës, parfois courtes et parfois rédigées par une machine.

Les descriptions comptent beaucoup de noms composés. Ils sont courants dans ce contexte parce qu'ils sont brefs et descriptifs. Traditionnellement, le traitement automatique du texte ignorait les noms composés. Comment utiliser des noms qui se suivent directement, sans relation explicite pour les relier ? Dans les requêtes généralement courtes et sans contexte, tous les mots sont importants. Il n'est pas question d'ignorer une paire de noms ou de ne considérer que le dernier nom d'un nom composé.

Nous entendons par un nom composé une structure composée de deux noms communs N_1 N_2 . Des structures comptant encore plus de noms existent, mais il est toujours possible de les décomposer en paires de noms. Certains noms composés sont définis dans le dictionnaire et leur sens a évolué loin de leur sens initial, comme *post office*. D'autres sont couramment utilisés. Certains n'existent que dans un jargon particulier où le sens des deux noms est évident. Enfin, certains noms composés sont générés sur-le-champ et quand même compréhensibles à la lumière du contexte, sans avoir à définir le terme préalablement. Parfois, ces deux noms appariés deviennent de nouveaux mots. Selon les noms utilisés, l'interprétation du nom composé révèle différentes relations. *Tear gas*, par exemple, cache une relation de causalité : *gas causes tears*. Les deux noms de *desert rat* sont reliés par une relation spatiale : *rat that lives in*

1. exemple fourni par l'entreprise 3CE

the desert (rat qui vit dans le désert). Comment, devant *love song*, déterminer que la relation est *chanson à propos de l'amour* et non une relation de causalité ou une relation spatiale ?

Dans le processus de création d'un nouveau nom composé, comme *olive oil*, la relation entre ces deux termes, *oil made from olives*, est perdue. Le défi est de retrouver cette relation. L'approche la plus rapide en traitement de langage naturel est d'ignorer les noms composés ou de les traiter avec une procédure très simple. Par exemple, ne considérer que le nom commun en dernière position suppose qu'il est toujours le plus important membre du nom composé, en espérant que le sens de la phrase soit préservé. Il est courant aussi, pour retrouver la relation cachée, de considérer le nombre de relations possibles comme limité. Ce faisant, le défi devient un problème de catégorisation : un nom composé doit avoir une relation parmi un certain nombre de possibilités.

Notre premier objectif est donc de découvrir la relation implicite liant les mots d'un nom composé donné en anglais. Dans le cadre des descriptions d'inventaire, où chaque mot est précieux, il est pertinent de prendre le temps de rendre explicite la nature et le sens de la relation au sein des noms composés. Elle pourrait fournir un indice de plus pour la classification d'un article dans le système harmonisé. Le système est basé en grande partie sur la relation de provenance et les noms des catégories SH contiennent eux-mêmes souvent des noms composés comme les *cocoa preparations* et les *fish eggs*. Trouver un nom composé avec une relation de provenance pourrait désigner le nom composé comme l'élément principal de la description d'inventaire. Le type de relation pourrait aussi influencer le choix final. De plus, selon le sujet et l'objet de la relation *syrup made from the starch of maize (corn)*, elle permettrait de déterminer si *corn syrup* doit être classé avec les produits de céréale ou avec les sucres et bonbons.

Nous posons l'hypothèse qu'il est possible d'identifier la relation au sein des noms composés à l'aide de corpus. Un corpus est une banque de textes utilisée comme source de contexte et d'information sur le monde en général. Un corpus nous permet de trouver des exemples contenant les deux noms. Nous pouvons nous servir du matériel extrait pour identifier la relation.

Nous avons choisi de baser notre méthode pour identifier la bonne relation sémantique sur l'approche de Nakov (Nakov and Hearst, 2008). La méthode de Nakov est facile à comprendre et sa ressource principale est le grand corpus de texte qu'est le web, accédé par le moteur de recherche Google. Toutefois, comme Google a restreint la possibilité de lui envoyer des requêtes en 2011, il n'est plus envisageable, pour notre système, d'accéder à ce corpus. Dans la version des termes d'utilisation de 2007, il est écrit :

You specifically agree not to access (or attempt to access) any of the Services

through any automated means (including use of scripts or web crawlers).²

Après 2012, le texte est changé à

For example, don't interfere with our Services or try to access them using a method other than the interface and the instructions that we provide.³

Malgré ce changement, nous allons reproduire, le mieux possible, l'approche de Nakov. Toutefois, il faudra trouver des corpus de remplacement.

Le second objectif de la recherche est de démontrer qu'on peut atteindre des résultats similaires à ceux de Nakov en utilisant d'autres corpus que Google. Plusieurs corpus ont le potentiel de répondre à ce besoin. Nous en avons sélectionné quatre. D'abord, nous employons comme corpus une version figée et indexée de tous les articles anglophones de Wikipédia, l'encyclopédie internet éditée par tous. Ensuite, nous essayons d'accéder à des phrases de l'internet à travers les morceaux de phrases de trois, quatre ou cinq mots récoltés par Google dans son recueil de n-grammes⁴. Pour remplacer Google, nous pouvons simplement utiliser un autre moteur de recherche pour accéder au même corpus que Nakov, le web, quoiqu'indexé différemment. Faroo⁵ est un moteur de recherche ouvert aux requêtes automatiquement générées, mais il est limité. Yahoo Boss, la façade de l'index web de Yahoo! pour les logiciels, devrait nous donner accès au web se rapprochant plus de Google.

Cette recherche se limite aux noms composés de la langue anglaise.

L'intérêt pour les structures particulières que sont les noms composés à la jonction du nom et du groupe nominal, remonte à plus d'un demi-siècle. Tout le contexte qui nous permet d'aborder la problématique des noms composés est expliqué au chapitre 2. Il existe déjà nombre d'approches qui catégorisent les noms composés autrement que celle de Nakov. Nous présenterons les forces et faiblesses des principales approches candidates, de notre point de vue, au chapitre 3.

2. <http://www.google.com/intl/en/policies/terms/archive/20070416/>

3. <http://www.google.com/intl/en/policies/terms/>

4. <http://storage.googleapis.com/books/ngrams/books/datasetv2.html>

5. <http://www.faroo.com/>

CHAPITRE 2 SÉMANTIQUE DES NOMS COMPOSÉS

Ce chapitre est la première partie de la revue de littérature, qui comprend aussi une brève revue des techniques pour l'interprétation des noms composés (chapitre 3) et une présentation de l'approche de Nakov (chapitre 4). Nous allons, dans ce chapitre, détailler les fondations nécessaires pour comprendre les approches des chapitres subséquents. Seront présentés d'abord la définition et les règles qui régissent les noms composés, puis les prédicats de Levi (Levi, 1978), qui peuvent décrire la relation sémantique au cœur des noms composés. Nakov réutilise les prédicats et les noms composés qui sont le fruit du travail de Judith Levi en 1978. Levi a été parmi les premiers auteurs en linguistique à examiner de près les noms composés.

2.1 Noms composés

Levi avance qu'il faut considérer un nom composé de plusieurs mots comme un seul nom et qu'il possède un lien plus fort que celui d'un groupe nominal. Levi s'intéresse aux noms comportant un modificateur suivi d'un nom principal, qu'elle appelle *noms complexes* (ou *complex nominals*). « Ni le nom principal ni le modificateur n'indiquent le sens du nom complexe au complet. » (voir Levi, 1978, p. 52) Plus de 30 ans plus tard, en 2013, Nastase et son équipe utilisent un terme pour désigner un nom principal avec possiblement des modificateurs : *base noun phrase* (syntagme nominal de base) (Nastase et al., 2013). Leur définition est plus générale, mais inclut les noms composés. Nous utiliserons donc le terme *nom composé* pour parler d'expression composée de deux noms communs, comme l'exemple de Nastase, *cacao tree*.

Pourquoi existe-t-il des noms composés ? Selon Levi, une première explication pour l'existence des noms composés est l'efficacité de la communication. Les noms composés fonctionnent comme une petite phrase qui est comprimée dans une autre phrase. La petite phrase est appelée paraphrase, parce qu'elle nous sert à clarifier le contexte d'un nom composé. La seconde explication est la facilité avec laquelle de nouveaux noms sont créés. Les nouveaux noms qui gagnent en popularité seront graduellement assimilés dans le langage. Les usagers de la langue forment de nouveaux noms composés presque instantanément. Les nouveaux noms composés sont acceptables même en étant inconnus, parce que leur signification peut être devinée sur-le-champ.

Les noms composés sont compris différemment selon le contexte. Levi donne l'exemple du contexte du conte *Alice au pays des merveilles*, où les fleurs peuvent parler. Ce contexte rend

acceptable un nom composé comme *floral comment*, parce que les fleurs partagent effectivement leur opinion par des commentaires. L'efficacité de ce processus pourrait expliquer les grandes variations dans la manière de former les noms complexes.

Certaines règles du processus de formation de nom composé de Levi doivent être respectées par tous les noms. D'abord, les noms composés doivent être généraux. Sa définition doit rester vraie pour toutes les choses désignées par chacun des deux noms. *Desert rat* ne pourrait pas désigner seulement **ce rat qui habite dans le Sahara**. Le nom composé peut désigner tous les rats qui habitent dans n'importe quel désert, et non pas un rat ou un désert particulier. Ensuite, les noms composés ne peuvent pas avoir de relation négative. *Salt lake* ne désignera pas les lacs sans sel. Enfin, les éléments des noms composés ne doivent pas être redondants : le nom composé *water lake* n'est pas accepté puisqu'un lac est par définition composé d'eau. Le nom modificateur doit identifier une particularité qui restreint le nom principal. Par exemple, dans *onion tears*, le modificateur *onion* précise le type de larmes, soient celles causées par des oignons coupés.

Levi considère qu'on peut imaginer qu'au départ, l'objet référé par le nom composé était désigné par une petite phrase imbriquée dans une plus grande phrase. Par exemple, elle suppose qu'avant l'avènement du nom composé *olive oil*, il ait fallu dire *Add some oil that comes from olives to the recipe*. La création du nom composé s'est produite en supprimant les mots *that comes from*. On peut supprimer des mots lors de la formation des noms composés parce que tous les mots supprimés font partie d'un ensemble fini et connu, toujours selon Levi. Cet ensemble contient toutes les relations sémantiques nécessaires. Chaque nom composé y trouve une relation suffisamment précise pour décrire le lien entre ses deux noms. Elle appelle ces possibilités des « recoverably deletable predicate » (voir Levi, 1978, p. 76), ou « prédicats qu'on peut supprimer et retrouver par la suite » (voir Levi, 1978, p. 59). Les douze prédicats, tels que FROM et ABOUT, sont le sujet de la prochaine section.

L'ambiguïté qui résulte de la suppression est limitée par le contexte et la nature des mots constituant le nom composé, du moins suffisamment pour permettre de comprendre le sens du nom composé sans interrompre la communication. Par exemple, interpréter *music talent* comme *talent in music* ou comme *talent for music* ne causera pas de difficulté pour la compréhension, même si la confusion reste possible. Il reste toujours une dose d'ambiguïté liée aux noms complexes, qui est acceptée et qui ne nuit pas en général. Pour une approche automatique, toutefois, trouver la bonne relation pour chaque nom composé quand il y a plusieurs relations possibles représente un défi de taille.

Le but premier de Levi est de démontrer que tous les noms composés sont formés par des paraphrases d'une façon logique et à portée d'entendement, même si leur définition évolue

avec le temps. Même si Levi fournit des classes pour les relations sémantiques, elle met en garde ses lecteurs contre la démarche de trouver un sens unique à un nom composé à partir de son usage usuel. Certains auteurs, comme Downing (Downing, 1977) et Zimmer (Zimmer, 1971), ont déjà soulevé des arguments en défaveur de la catégorisation des noms composés. Coulson (Coulson, 2001), plus récemment, croit que les relations au sein des noms composés sont trop riches en sens pour être énumérées. Elles sont le fruit de l'interaction entre des concepts complexes et uniques, selon Nastase. Dans la même lignée, Butnariu (Butnariu and Veale, 2008) propose un continuum de relations. Malgré ces oppositions, Ó Séahgdha (Ó Séahgdha, 2008), Vanderwende (Vanderwende, 1994), Rosario & Hearst (Rosario and Hearst, 2001), Lauer (Lauer, 1995) et Leonard (Leonard, 1984) ont tout comme Levi énoncé un ensemble fini de relations. En fait, on peut presque dire, selon Nastase, que chaque tentative d'analyser les relations génère une nouvelle liste de prédicats. Nous avons choisi d'utiliser l'ensemble de Levi parce qu'il se situe à un bon niveau d'abstraction et qu'il est utilisé par plusieurs auteurs, dont Nakov.

Certaines portions de la théorie de Levi sont trop importantes pour ne pas être mentionnées, mais resteront inutilisées dans le contexte de ce mémoire. D'abord, la théorie de Levi s'applique aussi aux adjectifs qui fonctionnent comme un nom modificateur dans le nom composé. Levi considère de la même manière les formes *urban life* et *city life*. *Urban life* n'a pas encore subi la dernière étape dans la formation des noms composés. Aucun de nos noms composés n'utilise d'adjectif, bien que Nastase et son équipe les permettent comme premier nom de leurs noms composés. Ensuite, nous éviterons complètement les noms complexes formés par le processus de *nominalisation*. Malheureusement, cette simplification entraîne le sacrifice des noms composés contenant un dérivé du verbe lui-même, plutôt que du sujet ou de l'objet du verbe. Par exemple, *Taxi driver* peut être exprimé comme X drives a taxi. Tous les noms composés seront plutôt traités par le premier processus, celui de suppression des prédicats, mais il y a là une belle source d'amélioration.

Nakov ne traite pas les noms composés qui ont perdu leur espace, comme *roommate*. La probabilité que leur définition se soit raffinée et généralement éloignée de la définition initiale est trop grande. En plus de ces omissions, il faut ajouter des exceptions supplémentaires et perdre quelques noms composés de plus, qui ne seront pas traités. Premièrement, les mots comme *ladyfinder* et *foxglove* sont exclus du discours de Levi parce qu'ils sont définis par l'emploi d'une métaphore. C'est plutôt difficile à traiter automatiquement. Deuxièmement, les noms qui désignent un élément par une partie de cet élément, comme *hammerhead* (le requin à tête de marteau) et *birdbrain* (une personne), ne peuvent pas être paraphrasés simplement et nous sont donc inaccessibles. Troisièmement, il y a les cas où l'élément principal et le modificateur sont égaux, par exemple dans *speaker-listener* et *player-coach*. Levi les a exclus

de sa démarche, mais Nakov a tenté de les catégoriser, notamment avec l'aide de certaines conjonctions de coordination. Dernièrement, il faut se limiter aux noms communs et ne pas traiter les noms propres composés comme *Sheridan Road* ou *President Obama*.

N'importe quel nom composé peut être incorporé comme élément modificateur ou principal dans un plus grand nom composé. Ce dernier est lui-même candidat à l'inclusion dans un nouveau nom composé. Le cycle peut se répéter théoriquement sans fin, récursivement. Il est responsable de noms composés géants tel que *industrial society recreational department financial manager psychological profile*, mais on trouve le plus souvent des noms composés de trois noms. Trois noms impliquent qu'il y a au sein de ces noms composés deux relations sémantiques internes, dont une fait le lien entre un nom et un nom composé. La démarche qui trouve la structure des noms composés de plusieurs noms se nomme le *bracketing*. Le *bracketing* fait la différence entre [**apple pie**] **plate**, l'assiette pour servir des tartes aux pommes, et **apple** [**pie plate**], l'assiette à tarte pour les pommes, beaucoup moins probable.

2.2 Prédicats

2.2.1 Prédicats pour paraphraser les noms composés

La section précédente décrivait la disparition de certains mots lors de la formation des noms composés. Tous les mots omis, retrouvables par déduction, peuvent être regroupés par type de relation sémantique sous des prédicats. Si l'on prend le nom composé général exprimé par $N_1 N_2$, la forme générale de sa paraphrase est N_2 PRÉDICAT N_1 , où N_2 est le sujet et N_1 est l'objet. La direction inverse est aussi possible : N_1 PRÉDICAT N_2 . Les prédicats sont souvent exprimés par un verbe, mais peuvent aussi être exprimés par une préposition, une conjonction ou potentiellement un adverbe. Prenons par exemple le nom composé *morning prayer*, dont la relation est exprimée par le prédicat IN. La préposition *during* suffit à représenter le prédicat dans la paraphrase *prayer during morning*. Tous les mots ou groupes de mots qui véhiculent l'idée d'un prédicat abstrait sont appelés des *éléments prédictifs*. Pour résumer, avec le nom composé *olive oil* paraphrasé par *oil comes from olives*, les éléments prédictifs sont *comes* ou *comes from* et dans l'extrait *oil from olives*, l'élément prédictif est la préposition *from*, seulement.

Selon Levi, les prédicats qu'elle propose sont suffisamment abstraits et suffisamment précis pour expliquer le processus de formation d'un nom composé et prédire le comportement d'un ensemble de noms composés. Nastase (Nastase et al., 2013) a la même opinion sur le niveau d'abstraction des prédicats : une unique relation sémantique pour tous les noms composés est

trop générale pour être utile et, à l'autre extrême, une relation par nom composé ne permet pas de généralisation.

Au niveau d'abstraction de Levi, il y a neuf prédicats : ABOUT, BE, CAUSE, FOR, FROM, HAVE, IN, MAKE et USE. Comme la paraphrase peut avoir deux directions, soit N_2 PRÉDICAT N_1 , soit N_1 PRÉDICAT N_2 , cette direction sera reflétée dans l'organisation des prédicats. Certains prédicats doivent être précisés. Si la phrase suit la forme N_2 PRÉDICAT N_1 , avec le second nom à la position du sujet et le premier nom à la position de l'objet du verbe, le prédicat sera noté PRÉDICAT₁. Le prédicat sera noté PRÉDICAT₂ dans la situation inverse, avec une phrase de la forme N_1 PRÉDICAT N_2 , c'est-à-dire quand l'objet est à la position du second nom et quand le sujet est à la position du premier. Bref, l'indice indique la position de l'objet de la phrase dans le nom composé.

En pratique, les prédicats CAUSE, HAVE et MAKE peuvent avoir la direction 2. La direction 2 des autres prédicats n'apparaît pas dans le langage ou est déjà représentée par un autre prédicat. Comme la direction 1 est plus commune, on peut écrire PRÉDICAT au lieu de PRÉDICAT₁. Le nombre total de prédicats passe à 12 : ABOUT, BE, CAUSE₁, CAUSE₂, FOR, FROM, HAVE₁, HAVE₂, IN, MAKE₁, MAKE₂ et USE. Chaque prédicat sera présenté à la prochaine section.

2.2.2 Prédicats proposés par Levi

En anglais, le second nom est à la position dominante du nom composé. Il est la tête du nom composé. Le premier nom est quant à lui appelé le modificateur. La relation ABOUT₍₁₎ établit un lien de sujet entre la tête du nom composé et le modificateur. La tête est souvent un acte de discussion ou le résultat de la discussion, souvent abstrait (*budget speech*). Le modificateur est beaucoup plus libre que pour les autres relations et n'a pour ainsi dire aucune contrainte sémantique. Les modificateurs de *tax law*, *love story* et *sports magazine* sont variés, mais leurs têtes ont toutes un rapport à la communication. Dans une phrase, la relation *magazine ABOUT sports* s'exprime comme : *A magazine concerning sports*.

La relation BE (ou BE₁) contient plusieurs sous-relations, mais on peut toujours formuler la paraphrase avec le verbe *is*. On trouve dans ce groupe principalement une relation de composition, comme dans *bronze statue*, une relation de sous-catégorie d'espèce, comme dans *flounder fish* et une relation métaphorique, comme dans *mother church*. La relation BE produit un grand nombre de noms composés avec une définition métaphorique. La relation de composition se recoupe grandement avec la relation MAKE₂. Elle est sensible à la direction, sauf dans les cas similaires à *king emperor*, *secretary-treasurer* ou *soldier ant* puisqu'on peut former avec *ant BE soldier* la phrase *ant that is a soldier by its role* tout comme la phrase *soldier that is an ant*. On peut former avec la relation BE le nom composé *house house* pour

désigner une maison typique, mais ce procédé n'est pas unique aux noms composés et les noms doublés ne sont que de très éphémères noms composés. La relation BE est la plus productive, avec 42 exemples annotés par Levi.

La relation CAUSE₁, contrairement à la relation précédente, contient peu de noms composés. La relation où l'objet est le premier nom se formule avec le suffixe *-causing*. Ainsi, pour *tear gas*, la paraphrase *gas CAUSE tear* peut être exprimé comme *tear-causing gas*. Au long, la définition est *the gas that causes tears*.

La relation CAUSE₂, qui est l'inverse de CAUSE₁, produit beaucoup plus de noms composés. En guise d'exemple, la phrase *Birth causes pain*, ou *birth CAUSE pain*, est une bonne paraphrase pour le nom composé *birth pain*. On peut conserver l'ordre du nom composé en l'exprimant par *birth-caused pains*.

La relation sémantique de but peut s'exprimer par le prédicat FOR. Ce prédicat est souvent présent dans les paraphrases par la préposition *for*, bien sûr. Prenons par exemple le nom composé *horse doctor* qui se dit au long *doctor for horses*. Le préfixe *-pro* est aussi utilisé parfois pour exprimer la relation au coeur de *abortion vote*, par exemple : *pro-abortion vote*. Il est difficile de choisir des verbes spécifiques pour les noms composés du prédicat FOR. Les noms composés *nose drops*, *bingo night* et *plant food* ne partagent que la relation de but.

Le prédicat FROM₍₁₎ peut représenter une relation de source naturelle ou de provenance. On représente *olive oil* comme *oil FROM olives* parce qu'il représente une source naturelle dans la phrase *oil that comes from olives*. FROM contient aussi des exemples tels que *sea breeze* dont la paraphrase est *This breeze comes from the sea*. Le premier nom est alors un lieu plutôt qu'une ressource.

Le prédicat HAVE₁ forme ses noms composés selon Levi en utilisant *have* ou *with* avec le premier nom comme objet pour établir une relation de possession. Pour trouver le verbe *have* dans le nom composé *apple cake*, il faut passer par la paraphrase *cake that has apple*. Le faible nombre de noms composés produit par ce prédicat peut être expliqué par la présence d'adjectifs qui peuvent déjà décrire la possession en précédant un nom, comme dans le cas de *wealthy person*.

Le prédicat HAVE₂ ressemble à HAVE₁, mais est beaucoup plus prolifique. Ceci est peut-être dû à la nature de la relation HAVE, qui accepte une grande variété de sujet et d'objet, tout comme le prédicat BE. Le nom composé *government land* est un bon exemple de la forme *government HAVE land* puisqu'il est facile de former la phrase *the government owns/uses/has this land*.

Un nom composé peut contenir une relation de localisation, qui sera désignée par IN₍₁₎. Ce prédicat est utilisé autant pour les localisations spatiales que temporelles. *Field mouse*, un

exemple de localisation spatiale, peut être défini par *Mouse that lives in the fields*. La relation temporelle se retrouve dans *childhood dreams*, qu'il est possible de voir comme *dreams IN childhood*. Il existe des rapports avec le lieu qui ne peuvent pas être rendu implicite, comme *between*, *not on* et *outside of*. Seule la prolifique relation de localisation peut être retrouvée après suppression.

Le prédicat **MAKE₁** exprime l'action de produire un objet physique, de créer le concept désigné par le premier nom de la paire. Il y a uniquement 4 paires de noms annotées par Levi pour cette relation, ce qui fait du prédicat **MAKE₁** le plus difficile à retrouver par système automatique. Cette relation est néanmoins acceptable pour Levi parce qu'elle se trouve au même niveau d'abstraction que les autres prédicats. Il est vrai que *honey bee* s'exprime correctement comme *bee MAKE honey* et par sa paraphrase *the bees that make honey*.

La relation **MAKE₂** ne ressemble pas comme on pourrait s'attendre à la relation précédente. Elle s'éloigne tant de la relation **MAKE₁** qu'elle éclipse en partie la relation **BE₍₁₎**. Elle exprime plus la composition que la création. L'exemple *daisy chains*, qui sont des chaînes où les mailles sont des fleurs, le démontre. Le nom composé peut être vu comme une composition de mailles, par la phrase *chains made of daisies* ou *daisies make chains*, tout comme il peut être vu comme une chaîne qui est aussi des fleurs, par la phrase *chains that are daisies by their nature*. Cet exemple illustre bien l'ambiguïté inhérente à la catégorisation des noms composés.

Finalement, le prédicat **USE₍₁₎** présente la relation d'utilisation d'un instrument. On la retrouve dans les groupes de mots à *l'aide de* et *grâce à*. Souvent, le second nom sera un engin et le premier nom sera une ressource qu'il utilise pour son bon fonctionnement. Par exemple, *steam iron* peut se définir au long comme *iron that uses steam to function*. On trouve aussi l'exemple plus classique *voice vote*, dont la structure *vote USE voice* se retrouve dans les phrases *Let the vote be by voice* et *A vote by means of voice*. Ici, aucun verbe précisant comment le vote utilise la voix n'est nécessaire.

De cette liste, Levi exclut la désuète relation **N₂ which reaches just to N₁**, qui a dans le passé formé des mots comme *waist coat*. Elle exclut aussi la relation **LIKE**, une relation d'analogie, de comparaison. **LIKE** est plutôt une caractéristique supplémentaire des noms composés dont la relation est plus métaphorique. Par exemple, *summer sunshine* se définit au long, *sunshine that is like a sunshine in summer*. La comparaison est présente, mais un autre des douze prédicats sera toujours présent, dans ce cas-ci, **IN**.

2.2.3 Vecteur prédictif

Les éléments prédictifs qui décrivent un nom composé peuvent être mis dans une liste pour décrire extensivement la relation implicite du nom composé. La liste ainsi formée s'appelle un vecteur prédictif. Le vecteur rassemble tous les éléments de contexte, qu'ils soient des verbes, des particules, des conjonctions ou plusieurs de ces types de mot, s'ils apparaissent ensemble dans la phrase d'origine. La validité de la description dépend de comment les éléments prédictifs sont isolés, de comment les phrases sont sélectionnées et de si la source des phrases, le corpus, est appropriée. Bien qu'en théorie tous les mots de la phrase peuvent fournir du contexte au nom composé, il est avantageux d'extraire les éléments prédictifs du segment de phrase coïncé entre les deux éléments d'un nom composé, à en croire Nastase (Nastase et al., 2013), Butnariu et Veale (Butnariu and Veale, 2008) et Turney & Littman (Turney, 2005). Pour finir, plus le vecteur prédictif est long, mieux sera décrite la relation de son nom composé, à condition qu'il ne soit pas pollué par des mots mal choisis.

CHAPITRE 3 APPROCHES POUR L'INTERPRÉTATION DES NOMS COMPOSÉS

Après avoir présenté le contexte des noms composés et de la relation à identifier, il est essentiel d'examiner les approches qui ont déjà abordé la problématique avec succès. Cette section présentera les travaux de Rosario, Girju, Moldova, Nastase, Kim et Nakov. Tous les travaux abordent différemment notre problématique. Nous choisissons de répliquer et de tenter d'améliorer l'approche de Nakov, jugée la plus utile et la plus simple, tout en atteignant de bons résultats.

Premièrement, Nastase et son équipe (Nastase et al., 2006) ont trouvé cinq relations sémantiques au sein des noms composés : **CAUSAL**, **PARTICIPANT**, **SPATIAL**, **TEMPORAL** et **QUALITY**. Ces catégories peuvent être raffinées en 30 relations sémantiques. L'apprentissage supervisé à la fin de la méthode est fait de trois manières différentes, avec trois engins : TiMBL, qui fait de l'apprentissage basé sur la mémoire, C5.0, qui bâtit des arbres décisionnels et SVMlight, une machine à vecteur de support. Deux approches distinctes sont essayées pour trouver la relation sémantique. Les deux commencent en apprenant un peu du sens de chaque nom individuel et par la suite devinent la relation sémantique. Les deux approches se distinguent par leur manière de trouver l'information qui sera donnée aux machines d'apprentissage. Une approche utilise les hyperonymes de Wordnet (Princeton University, 2010). L'autre interroge le corpus national britannique (BNC)¹ pour connaître les voisins immédiats des noms qui composent le nom composé inconnu. Si le nom est *child*, l'engin trouve des verbes adjacents comme *grow*, *eat*, *sleep* et *play*, ainsi que des adjectifs comme *good*, *happy*, *sad*, *small*, ... La méthode du corpus a l'avantage de ne pas avoir besoin de la désambiguïsation de sens de Wordnet, mais elle est fragile si le corpus est petit.

Bien que l'évaluation atteigne de très bons résultats, le problème tel que présenté n'a que cinq catégories sémantiques. L'approche atteint la limite des capacités d'information en entrée des systèmes d'apprentissage. Bien qu'elle nécessite beaucoup de ressources, il serait intéressant de l'essayer avec d'autres ensembles de relations sémantiques.

Deuxièmement, l'approche de Rosario et Hearst (Rosario and Hearst, 2001) se distingue parce qu'elle utilise une arborescence propre au monde de la médecine, le MeSH, dans laquelle le concept de *migraine* est représenté par C10.228.140.546.800.525. Chaque niveau est plus détaillé : *C* signifie maladie, *C10* signifie maladie du système nerveux, *C10.228* signifie maladie du système nerveux central, et ainsi de suite. Au total, 2245 noms composés ont été

1. <http://www.natcorp.ox.ac.uk/>

classés dans 18 catégories comme **Person afflicted**, **Instrument** et **Activity/Physical process**. L'algorithme de Rosario assigne à chaque nom composé une série de variables booléennes qui marquent la présence ou l'absence de noms individuels dans le nom composé. Comme l'expérience effectuée par Rosario comptait 1184 noms uniques parmi tous ceux qui apparaissent dans les noms composés, la série comptait 1184 variables concaténées. Un réseau neuronal à une seule couche cachée utilise les séries par la suite. Le système testé est d'abord entraîné avec la moitié des noms composés catégorisés, puis avec seulement 5% pour s'approcher des conditions normales d'analyse. Rosario atteint la très bonne performance de 61% pour le meilleur modèle, qui utilise 6 niveaux de profondeur de l'arborescence médicale et 52% en se limitant à une profondeur de 2 niveaux.

Son système performe bien quand se présentent des noms inconnus, particulièrement lorsque les noms inconnus sont à la seconde position, ce qui amène Rosario à supposer que le second nom est plus utile pour révélateur de la relation que le premier. Malheureusement, le système est limité au contexte de la médecine par sa dépendance au MeSH.

Le prochain système (Girju et al., 2009), celui de Girju & et coll., a atteint la première position pour la tâche de classification de relations sémantiques entre noms, en 2007, à l'atelier international de SemEval². Son système utilise, comme celui de Nastase, l'apprentissage supervisé pour apprendre à distinguer les catégories à partir d'indices variés sur le contexte et la nature des noms du nom composé. Toute cette information est récoltée grâce à de nombreux outils : NomLex-PLUS³, des dictionnaires remplis manuellement, un étiqueteur POS et un parseur syntaxique, un détecteur de rôle grammatical, Wordnet et deux systèmes complets de catégorisation sémantique. Chacun des systèmes fournit des éléments de réponse. La décision finale du système est prise en prenant en considération 18 critères allant de la position des noms dans la paire au rôle sémantique de la phrase autour du nom composé. Un des deux derniers systèmes de catégorisation est une implémentation du système de Moldovan (Moldovan et al., 2004) présenté un peu plus loin. Ce module choisit des catégories de Wordnet (Princeton University, 2010) de plus en plus spécifiques pour chaque nom des paires de noms composés. L'implémentation du système par Girju se souvient de tous ses résultats intermédiaires, quitte à prendre plus de temps et d'espace disque, pour choisir le résultat qui convient le mieux au nom composé à classer. Le second système de catégorisation sémantique est libSVM⁴, une machine à vecteur de support, utilisé pour choisir une relation à partir de multiples traits sémantiques. Girju a en plus ajouté aux 980 phrases de l'ensemble d'entraînement de SemEval plus de 6500 phrases supplémentaires, annotées par au moins

2. <http://nlp.cs.swarthmore.edu/semeval/>

3. <http://nlp.cs.nyu.edu/meyers/NomBank.html>

4. <http://www.csie.ntu.edu.tw/~cjlin/libsvm/>

deux juges.

Il est clair que le système de Girju est aussi compliqué qu’il est performant. Girju atteint une précision de 76% en classant sept catégories. Elle démontre que l’information contextuelle améliore grandement les performances du système. Elle démontre que ses améliorations au système de Moldovan, Semscat, donnent de meilleurs résultats. Cette performance démontre qu’utiliser les catégories de Wordnet fonctionne. Pour notre expérience, nous nous éloignons aussi du contexte de la compétition SemEval, où les phrases sont annotées, où le sens des noms est désambiguïsé et où il n’y a que sept catégories. Plutôt que de répliquer ce système, nous tentons d’obtenir des résultats comparables, mais en utilisant moins de composants et moins d’information connue d’avance. Nous espérons donner rapidement des résultats avec un système simple, tout en gardant en mémoire tout le potentiel de Wordnet.

Une quatrième manière de catégoriser les noms composés est aussi basée sur un dérivé de Wordnet. Les auteurs Kim et Baldwin (Kim and Baldwin, 2005) utilisent la mesure de similarité entre deux mots nommée `Wordnet::Similarity` (Patwardhan et al., 2003). Pour comparer deux noms composés, ils multiplient la similarité des premiers noms avec la similarité des seconds noms. Avec cette mesure, Kim et Baldwin peuvent établir des liens entre les noms composés d’un ensemble d’entraînement et les noms potentiellement inconnus des tests. Kim et Baldwin annotent leurs noms composés avec les 20 relations sémantiques de Barker & Szpakowicz (Barker and Szpakowicz, 1998) et atteignent le bon résultat de 53%. Leurs mesures de similarité leur permettent aussi de connaître l’importance relative du premier et du deuxième nom pour les résultats valides. Le second nom est le plus utile, mais les résultats atteignent un maximum de précision quand les deux éléments du nom composé ont la même contribution. L’approche pourrait gagner à avoir plus de noms composés connus, possiblement annotés automatiquement. Nous utiliserons les noms composés de Kim et Baldwin pour une évaluation à la section 6.2.2.

Cinquièmement, nous nous penchons sur le modèle de Moldovan et son équipe (Moldovan et al., 2004), qui ont bâti l’ensemble de relations sémantiques le plus volumineux avec 35 relations. Le contexte de la recherche permet les génitifs, noms composés avec nom ou adjectif et même les phrases adjectivales, même s’ils se concentrent comme nous sur les paires de noms. Pour source de texte, Moldovan a utilisé le texte des journaux de Wall Street fourni au concours TREC-9⁵ et les mots désambiguïsés de Wordnet. Comme il utilise un modèle d’apprentissage supervisé, les noms composés extraits ont été annotés par quatre juges. Ils avaient l’option d’indiquer que la définition du nom composé jugé avait trop changé pour que la relation sémantique initiale soit récupérable et de ne pas assigner de relation. Le

5. <http://trec.nist.gov/>

système bâti par Moldovan et al. exploite la classe sémantique de Wordnet 2.0 pour les deux noms. Les auteurs ont développé un modèle de dispersion sémantique qui raffine par itération la probabilité qu'un nom composé ait une relation étant donné l'information fournie par Wordnet.

Moldovan et al. ont noté que seules les paires de noms se transforment graduellement en nom unique et lexicalisé, ce qui confirme qu'il faut les traiter différemment. Moldovan, avec les mêmes paramètres que cette recherche, atteint le résultat moyen de 43%. Moldovan identifie comme source d'erreur la désambiguïsation automatique, des manques dans l'ensemble d'entraînement et le traitement combiné des noms composés issus de nominalisation ou non (voir la fin de la section 2.1).

Sixièmement et dernièrement, le système de Nakov et Hearst (Nakov, 2008) interroge le web, par l'intermédiaire de Google, pour trouver des informations sur le nom composé à catégoriser. Il trouve des verbes, des prépositions et des conjonctions qui apparaissent entre les deux noms du nom composé dans des phrases simples. Ces mots sont regroupés dans une liste et la liste est comparée à la liste de mots pour des noms composés dont la catégorie est connue. La catégorie donnée en résultat est celle de la liste de mots la plus similaire à celle des mots trouvés sur le web.

Ce système a l'avantage d'être modulaire et facile à modifier. L'analyse syntaxique qui identifie les verbes, prépositions et conjonctions peut être aisément remplacée par une autre manière d'obtenir ces types de mots. Le corpus est utilisé au début du processus uniquement, alors il est facile d'imaginer le remplacement de Google par un autre corpus. L'inconvénient majeur du système est justement qu'il utilise le corpus web par Google. Traiter les milliers de résultats d'internet nécessite du temps et depuis la fermeture de Google aux applications lançant des requêtes automatiquement, cette méthode doit utiliser un corpus différent. Le système est simple tout en étant performant. En 2007, il a obtenu le meilleur résultat à la tâche 4 de SemEval (Nakov and Hearst, 2007) pour son type de système. Il a classifié des noms composés parmi 7 catégories avec un taux de succès de 67%.

Nous choisissons de reproduire la méthode de Nakov telle qu'elle a été rapportée dans son article de 2008. Cette méthode est décrite en détail au chapitre 4.

CHAPITRE 4 TECHNIQUE D'EXTRACTION DE LA RELATION DE NAKOV

La méthode de Nakov étant centrale à notre mémoire, nous y consacrons un chapitre complet pour l'expliquer étape par étape. Après une description générale, la structure de ce chapitre suit l'ordre des étapes de l'approche. Nous présentons d'abord l'extraction du corpus, puis l'application des patrons. Nous expliquons ensuite les vecteurs de référence et l'étape de comparaison, qui utilise les vecteurs de référence. Après, nous abordons la façon de déduire la relation sémantique correcte à partir des résultats précédents. La dernière section décrit la façon d'évaluer l'approche.

4.1 Description générale

Nakov (Nakov, 2013) a testé son approche avec les relations de Levi (Levi, 1978), les relations de Nastase (Nastase and Szpakowicz, 2003), les relations de Kim (Kim and Baldwin, 2006), les relations de Barker (Barker and Szpakowicz, 1998) et avec les noms composés de la tâche de la conférence de SemEval en 2007 (Girju, 2007). Les théories de Levi en particulier sont très importantes pour la syntaxe et la sémantique des noms composés, selon Nakov. Parmi les 250 noms composés, rassemblés en catégories, que Levi a décrits en annexe de son livre de 1978 (Levi, 1978), seulement 214 noms composés ont été utiles : ceux qui sont obtenus par suppression de prédicat.

Le but de l'approche est de trouver la relation sémantique qui relie les deux noms d'un nom composé donné. Comme aucun contexte n'est donné, des éléments prédictifs (verbes, prépositions et conjonctions) sont extraits de l'immense corpus qu'est le web, par l'entremise de Google.

En considérant chaque mot comme une dimension dans un espace vectoriel, on obtient un vecteur, appelé *vecteur prédictif*. On construit aussi un vecteur pour chacun des 214 noms composés de Levi. Ces derniers servent de référence. Le vecteur du nom composé inconnu est comparé à chacun des vecteurs de référence, et on lui attribue la relation du nom composé dont le vecteur est le plus proche (ou le plus similaire).

Les noms composés connus sont utilisés à la fois pour déduire la relation et pour évaluer la performance par l'approche Leave-one-out.

L'illustration 4.1 présente les modules principaux de la méthode de Nakov. « Tear gas » est le nom composé à l'entrée du système. Le premier module fait l'extraction de milliers de

documents de contexte à partir de l'Internet, et autant de phrases dans lesquelles apparaissent les noms principaux : *tear* et *gas*. Ces phrases sont ensuite filtrées par des patrons, dans le module d'analyse. Ce module, les patrons qu'il faut lui fournir et sa sortie, un vecteur prédictif, sont décrits à la section 4.3. Les vecteurs prédictifs de référence (section 4.4) sont utilisés à l'étape de comparaison, qui trie les vecteurs de référence en ordre décroissant de similarité. À partir des vecteurs de référence triés du module précédent, il est très rapide de prendre une décision sur la catégorie du nom composé « *tear gas* ». On choisit tout simplement la catégorie du vecteur le plus similaire. La sortie du système pour cet exemple est la relation de Levi CAUSE₁, qui est le résultat correct.

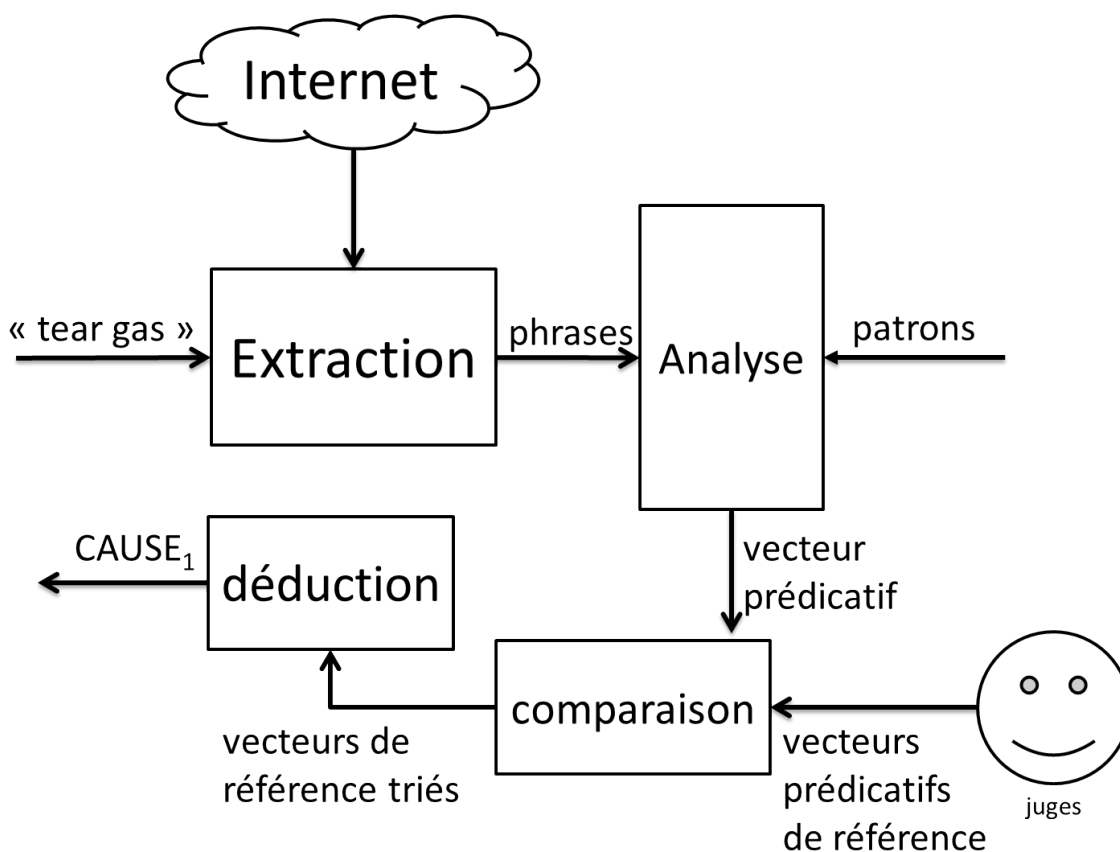


Figure 4.1 Méthode de Nakov de base

4.2 Extraction du corpus

Cette section explique l'utilisation du moteur de recherche Google. Elle explique aussi comment est fait le traitement des textes retenus pour en faire des phrases suffisamment appropriées pour l'analyse à l'étape suivante.

Des pages de résultats, seuls les courts textes, les *snippets*, fournis par Google pour chacun des sites indexés sont utilisés. Les sites web des résultats ne sont pas visités un à un. Jusqu'à 1000 résultats sont utilisés par requête. Chaque résumé correspond à un « document ». Plusieurs requêtes sont émises par nom composé pour aller chercher plus de résultats. Des variations dans la façon de formuler la structure **NOM * NOM** entraînent plus d'une centaine de requêtes, si toutes les combinaisons d'opérateurs et d'inflexion sont essayées. Voici la forme générale des requêtes :

$$infl_1 THAT * infl_2 \quad (4.1)$$

$$infl_2 THAT * infl_1 \quad (4.2)$$

$$infl_1 * infl_2 \quad (4.3)$$

$$infl_2 * infl_1 \quad (4.4)$$

Les premier et deuxième noms du nom composé, leurs formes au singulier ou au pluriel et leurs synonymes occupent la position $infl_1$ et $infl_2$, respectivement. Le mot-clé **THAT**, qui peut être présent ou non, est utilisé dans le même but : extraire plus de phrases du corpus de Google. **THAT** peut représenter à la fois les mots *that*, *which* et *who*. Il peut y avoir jusqu'à huit astérisques de suite dans la requête à Google.

Il est à noter que plusieurs documents seront extraits plusieurs fois, par différentes requêtes au même corpus. Devant cette réalité, les doublons sont filtrés pour ne pas augmenter indûment l'influence des phrases tirées de ces documents. On filtre les phrases qui ne contiennent pas les deux noms du nom composé, N_1 et N_2 , et on s'assure que les phrases candidates contiennent des mots entre N_1 et N_2 , puisque les éléments prédictifs d'intérêt se trouvent là. Le lemmatiseur Morphy de la suite OpenNLP¹ est utilisé pour savoir si le mot qui suit immédiatement l'instance du patron détecté dans le corpus est autre chose qu'un nom. Ceci indique que la phrase est complète et bien rapportée par Google, ce qui devrait augmenter la qualité des éléments prédictifs extraits.

Google rapporte peu de résultats pour certains noms composés. Pour augmenter le nombre de résultats, les noms du nom composé sont remplacés par un synonyme. *Abortion problem* est remplacé par *abortion trouble*, par *miscarriage problem* et même par *miscarriage trouble*. C'est Wordnet qui fournit les synonymes (voir la section 5.2.3 pour plus de détails). Ces substitutions ne sont valides que parce que le lien sémantique dans le nom composé n'est pas affecté par le remplacement d'un nom par son synonyme. Ainsi, la relation au sein de

1. <https://opennlp.apache.org/>

home cure et de *home remedy* devrait être la même, si *remedy* est un mot sémantiquement proche de *cure*. C’est une hypothèse qui permet d’augmenter le nombre de phrases extraites du corpus et par là, d’augmenter aussi le nombre d’éléments prédicatifs révélateurs de la relation.

4.3 Analyse

Après les requêtes au corpus et l’extraction des phrases individuelles, il est maintenant temps d’expliquer l’analyse syntaxique de surface et les patrons qui filtrent les phrases. Il y a différents patrons pour différentes phrases, qu’elles aient la forme « sujet verbe complément » ou qu’elles contiennent une clause relative, une particule, des verbes au passif ou des groupes nominaux particuliers. Les patrons permettent d’identifier un élément prédicatif, que ce soit un verbe (*make*), un verbe avec particules (*made of*), une préposition (*in*) ou une conjonction de coordination (*or*). Tous ces éléments sont des éléments prédicatifs.

Les patrons, en gros, sont des filtres pour du texte. Ils s’écrivent comme une phrase, avec des mots écrits normalement, qui doivent être retrouvés en ordre dans la phrase candidate. Ils peuvent aussi spécifier qu’une position peut contenir un ou plusieurs mots inconnus et ignorés, ou spécifier qu’un mot inconnu à un certain endroit doit être extrait de la phrase et présenté en sortie du patron. Un patron pourrait par exemple trouver tous les adjectifs qui précèdent le terme spécifique *guitar*, tout en faisant partie du même groupe nominal. Selon Nastase (Nastase et al., 2013), les patrons ont un faible taux de rappel, mais une haute précision. Ils ne capturent pas beaucoup de résultats parce qu’ils ne peuvent capturer que les phrases qui correspondent exactement aux filtres, souvent peu nombreuses. Ils permettent cependant une haute précision parce que si une phrase est reconnue par un patron suffisamment spécifique, elle remplit assurément toutes les conditions du patron, ce qui assure généralement un bon résultat.

D’abord, l’approche de Nakov extrait les verbes, sous forme lemmatisée et obtient leur nature grammaticale, représentée par les étiquettes morphosyntaxiques de PennTree (Santorini, 1990). Pour faciliter le travail de l’étiqueteur et diminuer ses erreurs, le segment de phrase *We look at the* remplace le début des phrases avant la structure **NOM₁ that PRÉDICAT NOM₂**. Il s’agit d’un ajout à la méthode fait après l’article de 2008 (Nakov and Hearst, 2013). Les verbes sont repérés par l’étiquette **VB**. Ils doivent avoir un des deux noms-clés comme sujet et l’autre nom-clé comme objet. Par exemple, « The *committee* includes many *members* », dans lequel le verbe souligné est l’élément extrait.

Il y a beaucoup de structures de phrase qui ont le même sens. Plus le nombre de patrons

est grand, plus il est possible de tirer avantage de différentes structures de phrase, ce qui se traduit par un plus grand nombre d'éléments prédicatifs capturés. Les choix des patrons ont un grand impact sur la quantité de verbes extraits, mais aussi sur leur pertinence dans l'identification de relation. Les patrons de Nakov couvrent 5 grandes structures de phrases.

- Les premiers patrons détectent les phrases simples de la forme *sujet prédicat complément* : *Five members constitute a committee.*
- Les patrons n'acceptent une phrase que si les deux noms sont à la fin de leur groupe nominal, dans une position dominante : *Senior members constitute a city committee.*
- Les éléments prédicatifs sont aussi extraits quand un premier nom est décrit par une clause relative qui abrite un verbe et son complément, par exemple dans la phrase *members that make up the committee.*
- Un type de patrons est nécessaire pour extraire les verbes accompagnés de particules, comme *make up* dans le même exemple : *members that make up the committee.*
- Les verbes à la forme passive sont couverts par leur propre version des patrons précédents. *The committee is made up of five members.*

Les résultats de l'analyse syntaxique de surface et de l'application des patrons sont les éléments prédicatifs. La direction de la phrase de laquelle ils sont extraits, $N_1 \dots N_2$ ou $N_2 \dots N_1$ pourrait être gardée en mémoire. Cette information est mentionnée par Nakov, mais ne semble pas être utilisée dans la suite de l'expérience. Une variation de la démarche de Nakov utilisant cette information sera présentée au chapitre 5.

Les résultats ne comptent pas les auxiliaires de verbe *will*, *can*, *may*, etc. Les verbes sont toujours lemmatisés, sauf si le verbe est au passif. Dans ce cas, le mot *be* précède le verbe conjugué au passé. Les verbes sont importants, mais les prépositions et les conjonctions sont aussi extraites. Par exemple, la préposition *in* est extraite de phrases comme *member in the committee* et la conjonction de coordination *and* est extraite de la phrase *Sam is a coach and a player*, pour le nom composé *player coach*. Les éléments prédicatifs sont regroupés dans une liste pour donner un vecteur prédicatif.

4.4 Vecteurs prédicatifs de référence

Une fois qu'un vecteur prédicatif a été assemblé pour décrire un nom composé quelconque, il reste à utiliser cette information pour deviner quelle relation sémantique sera attribuée au nom composé. Pour chaque nom composé fourni en exemple par Levi, un vecteur prédicatif est assemblé. Ce sont les vecteurs de référence. Ces vecteurs pourront être comparés au vecteur prédicatif du nom composé inconnu. Le nom composé sera alors jumelé à celui des noms composés de référence qui aura le vecteur qui ressemble le plus à celui du nom inconnu.

On supposera alors que la relation sémantique est la même pour les deux noms composés, ce qui nous permettra d’atteindre notre but initial : assigner une relation au nom composé inconnu.

Les vecteurs prédicatifs de référence sont importants et doivent être de qualité. Des juges humains ont donc été recrutés pour constituer ces vecteurs. Ils ont été recrutés par l’intermédiaire de Mechanical Turk, un service d’Amazon. Un total de 174 juges ont contribué des paraphrases en remplissant un formulaire.

Les instructions du formulaire sont

You are asked to substitute the dots with one or more verbs optionally followed by a preposition : “noun1 noun2” is a “noun2 that noun1”.

Le formulaire compte plusieurs noms composés. Pour chacun, le formulaire fournit une phrase à compléter, avec N_1 et N_2 positionnés, comme dans la phrase « “smoke signals” are signals that smoke(s) ». La phrase à compléter est toujours orientée dans le même sens.

On a demandé aux juges un minimum de trois éléments prédicatifs par nom composé. Un minimum de 25 juges différents a fourni des termes pour chaque nom composé. La moyenne d’éléments prédicatifs ainsi accumulés est de 71 éléments par nom composé.

À cause de la formulation des instructions aux juges, tous les éléments prédicatifs s’insèrent dans des phrases dont l’orientation est N_2 **prédicat** N_1 , où le deuxième nom précède l’élément prédicatif et où le premier nom vient après.

Les éléments prédicatifs de référence peuvent ne pas être strictement un verbe, mais un verbe suivi d’une particule comme l’élément *make of* pour le nom composé *committee member*, tiré de la phrase « committee made of members ».

L’approche de Nakov lemmatise les verbes parmi les éléments prédicatifs, sauf s’ils sont à la forme passive. Dans ce dernier cas, jusqu’à trois mots constitueront l’élément prédicatif : *be*, le verbe au passé et la particule qui suit en général, comme dans le cas de *be made of*.

Demander aux juges trois éléments peut générer des éléments de moindre qualité, particulièrement en troisième position. Heureusement, l’ordre des termes est connu, ce qui permet de filtrer a posteriori les réponses des juges. Un test a été fait pour savoir si l’ordre des réponses sur le formulaire est indicateur de leur qualité. L’algorithme de Nakov a été exécuté d’abord en utilisant l’ensemble des éléments et ensuite uniquement en utilisant les éléments qui étaient en première place dans les formulaires (933 différents termes). L’approche qui utilise tous les 2695 termes a été légèrement meilleure, mais pas significativement. Les relations les moins ambiguës ont vu leur score augmenter, mais l’inverse s’est produit pour les relations qui ont moins d’exemples : leur score s’est détérioré. Uniquement afin de comparer les

deux approches, leur similarité cosinus a été mesurée (Nakov and Hearst, 2013). La similarité cosinus est presque identique pour tous les éléments et pour les premiers éléments du formulaire seulement, respectivement de 32% et 31%. Ces deux pourcentages sont trop près pour pouvoir prendre une décision. Les deux approches sont donc interchangeables pour l'étape de comparaison. Nous testerons les deux lorsque nous répliquerons l'approche complète (voir 6.3.4).

4.5 Déduction de la relation

Cette section décrit le processus de décision qui part des noms composés de référence triés et qui fait le choix de la relation sémantique la plus probable pour le nom composé inconnu. Armé des vecteurs prédictifs de référence où chaque élément apparaît un certain nombre de fois, on établit le score TF-IDF, plus utile que les fréquences, pour chacun des éléments. On chiffre la similarité entre le vecteur qui sert à identifier le nom composé inconnu et chaque vecteur des noms composés de référence grâce au coefficient de Dice (défini plus loin). Il suffit ensuite de choisir la relation sémantique du nom composé le plus ressemblant, par l'approche du plus proche voisin.

Tout d'abord, la fréquence d'apparition des éléments communs à chaque vecteur prédictif a été transformée en une mesure plus utile, le TF-IDF (« Term Frequency - Inverse Document Frequency »). Utiliser la mesure TF-IDF diminue l'importance des éléments courants et augmente l'importance relative des termes qui sont utilisés dans quelques rares circonstances et dont la présence est révélatrice d'une relation sémantique unique. On souhaite diminuer l'impact des verbes abondants tels que *be*, *have* et *do*. À l'inverse, on veut augmenter l'impact des verbes rares, tels que *sanction*, *authorize*. On calcule la valeur de TF-IDF de chaque élément prédictif avec la formule 4.5, dans laquelle TF est la fréquence de l'élément dans le vecteur et N est le nombre de noms composés de référence plus un pour le nom composé inconnu.

$$TF-IDF = TF \cdot \log \frac{N}{DF} \quad (4.5)$$

Le DF de l'équation signifie *document frequency* (*fréquence dans les documents*). Pour notre calcul, chaque vecteur prédictif est considéré comme un document. DF est donc le nombre de vecteurs qui contiennent l'élément prédictif, incluant autant les vecteurs de référence que le vecteur prédictif en provenance du corpus.

Le coefficient Dice (Dice, 1945), décrit à la formule 4.6, est la mesure qui est utilisée pour mesurer la similarité de deux vecteurs, D et Q . Chaque élément des vecteurs est considéré

indépendamment. Seuls les éléments identiques contribuent à augmenter le coefficient.

$$DICE = 2 \cdot \frac{|D \cap Q|}{|D| + |Q|} \quad (4.6)$$

La formule générale a été adaptée au besoin de la méthode, pour donner la formule 4.7. Dans l'équation, un a est un score TF-IDF pour le premier vecteur tandis que b représente un score pour le second vecteur.

$$Dice(A, B) = \frac{2 \cdot \sum_{i=1}^n \min(a_i, b_i)}{\sum_{i=1}^n a_i + \sum_{i=1}^n b_i} \quad (4.7)$$

Le plus petit des deux scores, $\min(a_i, b_i)$ est utilisé pour éviter qu'un élément qui n'a qu'une faible fréquence influence grandement le résultat parce que l'élément apparaît souvent dans l'autre vecteur. Si un élément apparaît très peu souvent dans un vecteur, son poids sera petit, même si sa fréquence est très élevée dans l'autre vecteur. Le calcul ignore tous les éléments qui ne sont pas présents dans les deux vecteurs à la fois. D'autres approches, comme la similarité cosinus, auraient pu remplir le rôle du coefficient Dice.

Selon Nastase (Nastase et al., 2013), il est important de normaliser les résultats quand on utilise le web, pour que la popularité d'un mot sur la toile n'affecte pas la déduction. Tous les résultats sont normalisés pour que les coefficients Dice soient entre 0 et 1. La relation de Levi sélectionnée est celle du nom composé le plus similaire au nom composé inconnu. Le plus proche voisin est simplement celui dont le vecteur prédictif a obtenu le score Dice maximum.

4.6 Évaluation

L'approche de validation croisée Leave-one-out est utilisée pour évaluer les résultats. L'approche retire de l'ensemble de référence une paire de noms et utilise cette paire de noms comme si sa relation sémantique était inconnue. Elle permet de comparer le résultat obtenu et le résultat connu initialement pour savoir si la méthode a réussi ou échoué. La procédure répétée sur toutes les paires de noms de l'ensemble d'apprentissage permet d'avoir un ensemble de tests équivalent en taille à l'ensemble d'apprentissage. L'ensemble d'apprentissage correspond aux 214 noms composés donnés en exemple par Levi.

Cette approche a toutefois l'inconvénient majeur de refléter principalement la cohérence des exemples de référence les uns par rapport aux autres et ne révèle pas nécessairement le comportement de l'algorithme devant des cas complètement nouveaux. L'approche induit des erreurs si l'élément retiré est le seul exemple d'une situation particulière ou s'il y a très peu d'exemples similaires, ce qui est le cas pour la relation **Make**₁, représentée par seulement

quatre noms composés. Pour cette relation, il faut juger si le vecteur prédicatif est similaire à celui d'un des trois noms composés restants. C'est trop peu.

La limite inférieure du test, que tous les algorithmes doivent dépasser est l'algorithme de la classe majoritaire. La classe majoritaire est la catégorie la plus représentée dans les exemples de Levi. C'est la relation BE avec 42 exemples. Une machine qui supposerait que la relation inconnue est toujours BE atteindrait le bon résultat 42 fois sur 214 essais. La limite inférieure du test est donc d'environ 20%.

L'approche de Nakov a aussi une limite supérieure, selon Nakov. Il s'agit du pourcentage de succès si les éléments prédicatifs extraits du web réussissent à être similaires en qualité et en nombre à ceux donnés par les juges humains. La pensée ici est que l'algorithme qui réussit à extraire d'un corpus le vecteur idéal trouvera les relations sémantiques implicites avec le meilleur taux de succès. Comme le vecteur prédicatif idéal est meilleur ou de même qualité que les vecteurs des juges humains, la limite supérieure est établie en remplaçant le vecteur prédicatif du corpus par le vecteur de référence, pour un nom composé connu. Or, l'approche *Leave-one-out* utilisée ne permet pas un résultat théorique de 100% au test. Quand les vecteurs des juges humains sont donnés au classificateur, chaque nom composé est comparé aux autres exemples de Levi. Cette stratégie d'évaluation donne un résultat maximum de 78% pour tous les éléments prédicatifs et de 72% seulement lorsque seuls les premiers éléments des formulaires sont utilisés.

Le meilleur taux de succès de l'approche est de 50%. Il est obtenu en utilisant les verbes et les prépositions seulement. Ajouter ou non les conjonctions de coordination n'a aucun impact. Ajouter le mot-clé **THAT** pour la recherche de phrases dans Google ne nuit pas, mais n'augmente pas les résultats de façon significative. Le pire résultat est de 32%. Il est obtenu en n'utilisant que les prépositions et les conjonctions de coordination. L'utilisation des conjonctions comme éléments prédicatifs n'aide pas. Les verbes sont les éléments prédicatifs qui ont un impact réel sur les résultats. Les méthodes qui n'utilisent que les verbes atteignent 46%, avec le mot-clé **THAT**. Les verbes seuls égalent presque, considérant l'incertitude, le meilleur résultat. Ce qui laisse à penser que « dans l'ensemble, les plus importants éléments de phrase sont les verbes » (traduction libre) (Nakov and Hearst, 2013).

Ce résultat ne peut pas être comparé à la tâche de classification de nom composé (tâche 4) de la conférence de SemEval en 2007 (Girju et al., 2007), où les résultats dépassent ou atteignent 50% de succès. Il y avait sept plutôt que douze catégories, mais surtout, chacune des catégories est traitée indépendamment, avec une réponse binaire. Soit une relation donnée existe pour un nom composé, soit non. De plus, le sens Wordnet était fourni, ainsi qu'une phrase de contexte.

Le résultat de Nakov peut être comparé à Rosario (Rosario and Hearst, 2001) et à Kim (Kim and Baldwin, 2005). Rosario, dans un environnement contrôlé, le dépasse avec un taux de succès de 67%. L'approche de Nakov est légèrement moins fructueuse que celle de Kim, à 53%, qui choisit la bonne relation entre 20 catégories.

Dans l'ensemble, selon Nakov, il y a une corrélation positive entre le nombre d'éléments prédicatifs extraits des corpus et les résultats. Il faudrait effectuer le plus de requêtes web possible. C'est pourquoi le web, par sa taille, est le meilleur corpus. Toutefois, Nakov observe que « l'utilisation des vecteurs proposés par les humains atteint une très haute précision tout en utilisant sept fois moins d'éléments prédicatifs » (traduction libre) (Nakov and Hearst, 2013). La qualité des éléments prédicatifs comme leur nombre a un impact positif sur le taux de succès de l'expérience. Que les meilleurs résultats soient actuellement atteints avec le web ne change pas le potentiel d'amélioration d'une meilleure sélection d'éléments prédicatifs.

Certains noms composés, dits *exocentriques*, doivent être définis en utilisant des définitions complexes, qui s'étendent sur plus d'une phrase. La structure de la définition s'éloigne de la simple structure **Nom Prédicat Nom**. *Monkey wrench* et *flea market* (Levi, 1978) sont des exemples de noms composés exocentriques, qui ne peuvent pas être décrit par l'approche de Nakov. Pour décrire *monkey wrench*, il faut faire référence au terme marin pour une petite pièce d'équipement, *monkey* et expliquer l'évolution du terme au courant des deux derniers siècles.

Par ailleurs, toujours selon Nakov, les noms composés qui ont pour structure un nom suivi d'un identificateur ou d'un nombre sont problématiques, comme *vitamin D*, *route 66* et *interferon alpha*. Les noms composés importés du français causent aussi certaines difficultés parce que le nom dominant est celui de gauche : *planet Earth* et *President Obama*. Finalement, une faille majeure de l'approche est que le succès individuel de la catégorisation d'un nom composé dépend de sa fréquence d'utilisation sur internet. *Chop suey*, par exemple, est mal identifié parce que *suey* est un mot très rarement utilisé en comparaison aux mots *love* et *song* du nom composé *love song*. Nakov propose de continuer à utiliser les synonymes, voire même les hyponymes ou les cohyponymes suffisamment similaires aux noms du nom composé.

CHAPITRE 5 MÉTHODOLOGIE

5.1 Description générale

Dans ce chapitre, la démarche que nous appliquons pour répliquer et enrichir l’approche de Nakov sera expliquée du début à la fin. Ce faisant, nous atteignons deux objectifs. D’abord, nous confirmons l’efficacité de la méthode de Nakov. Ensuite, nous validons notre hypothèse principale, soit qu’il est possible d’identifier la relation sémantique qui lie les deux noms des noms composés sans l’utilisation d’Internet. Nous expliquerons la démarche par laquelle l’expérience de Nakov est répliquée. La méthodologie adoptée suit la même structure que l’approche de Nakov illustrée à l’image 4.1 du chapitre précédent. Nous avons choisi cette structure pour souligner que le but premier de cette démarche est de confirmer l’expérience de Nakov en la reproduisant. Similairement, notre implémentation de l’algorithme commence par trouver dans un corpus des phrases où les deux noms du nom composé apparaissent ensemble. Puis, ces phrases sont analysées pour en extraire des éléments prédicatifs. Ces éléments prédicatifs sont ensuite comparés à des éléments prédicatifs de référence, obtenus grâce à l’aide de juges humains. Les comparaisons révèlent laquelle des relations sémantiques est donnée en réponse, à la fin de démarche.

Nous ne faisons pas que reproduire l’expérience. Nous expliquerons aussi l’ensemble des variations faites pour améliorer les résultats, parfois avec succès. Des variations sont introduites à tous les points du protocole : dans le choix des corpus, dans l’ajout de synonymes, dans le choix des patrons d’extraction et des vecteurs de référence et dans la manière de choisir la relation sémantique donnée en réponse.

5.2 Extraction du corpus

Cette section décrit la démarche utilisée pour le premier module de l’expérience. Il prend un nom composé dont la relation sémantique interne est inconnue et il trouve des phrases permettant de comprendre la relation entre les deux noms de la paire. L’image 5.1 montre les sous-étapes de l’extraction du corpus, du nom composé en entrée aux phrases candidates. L’extraction du texte dans le corpus telle quelle est centrale, mais elle n’est pas la seule étape. L’influence des synonymes pour former des noms composés équivalents sera testée pour savoir si elle aide les noms composés sous-représentés dans le corpus. Les textes du corpus ne peuvent pas être envoyés directement à l’analyseur syntaxique du module suivant. Il faut éliminer les textes identiques et séparer et filtrer les phrases dans l’étape de préparation. Commençons

en décrivant des corpus qui peuvent être inclus dans l'expérience.

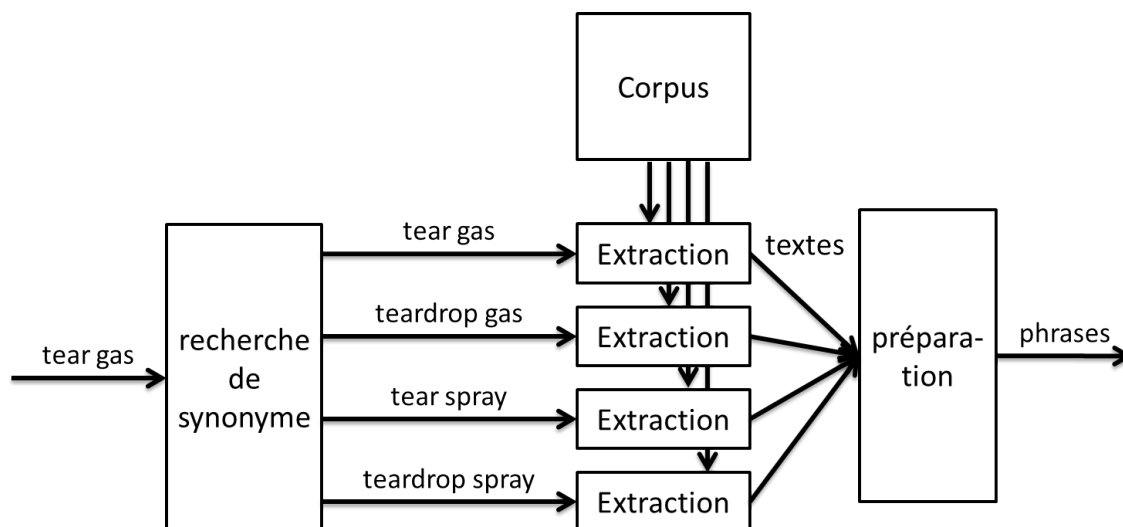


Figure 5.1 Extraction de contexte à partir du corpus

5.2.1 Description des corpus utilisés

Pour trouver des phrases fournissant du contexte aux noms composés, nous n'utiliserons pas le moteur de recherche Google puisqu'il a fermé l'accès aux requêtes automatisées. Il nous est donc impossible de répliquer en 2014 l'approche de Nakov complète telle que décrite en 2009. Pour comparer nos résultats aux siens, nous devons utiliser ses résultats intermédiaires, qu'il a rendus publics. Les éléments prédicatifs récoltés grâce aux résultats de Google seront utilisés pour déduire la relation sémantique, mais sans pouvoir tester les requêtes ni les patrons pour recueillir les éléments prédicatifs. Nous avons tout de même essayé des variations de l'expérience de Nakov en utilisant d'autres corpus : Yahoo!, les n-grammes de Google, Wikipédia et Faroo.

Yahoo! est le moteur de recherche qui remplace le mieux celui utilisé initialement dans l'approche de Nakov. Il fonctionne de la même manière. Nous utilisons Yahoo Boss pour faire des requêtes automatisées en Json. Nous avons développé un total de 22 445 requêtes qui ont donné 1 027 000 pages pour les 212 exemples de Levi avec leurs synonymes. Nous avons limité à 1 000 le nombre de pages retournées pour chaque nom composé. Comme Yahoo traite de la même manière les mots au singulier et au pluriel, nous utilisons la paire de noms initiale sans mettre au singulier ou pluriel N_1 ou N_2 . Par exemple, à partir du nom composé *desert rat*, nous envoyons la requête `http://yboss.yahooapis.com/ysearch/web?q=+desert+rat&format=UTF-8&start=0&abstract=long` .

Pour l’encodage de la requête, nous avons mis des symboles + devant les noms pour n’obtenir que les résultats qui contiennent à la fois *desert* et *rat*. Yahoo ne permet pas de spécifier qu’il faut un mot entre les deux noms de la requête. La valeur **start** demande à Yahoo! les 50 résultats à partir de ce chiffre. Il faut donc 20 requêtes pour obtenir les 1000 résultats au maximum. Si la page vers laquelle pointe Yahoo! est inaccessible, nous utilisons le texte du résumé mis en cache par Yahoo!, que nous obtenons avec la variable **abstract**. Nous accédons aux liens individuellement et transformons les pages obtenues, de format HTML, en texte avec l’aide de Jsoup¹. Avec *desert rat*, nous obtenons 534 résultats.

Cette approche est différente de l’approche de Nakov parce que nous allons sur les pages complètes quand elles sont disponibles, alors que Nakov se limitait au texte du résumé. C’est une manière d’augmenter le nombre de phrases de contexte trouvées.

Un deuxième corpus que nous essayons pour remplacer Google, c’est les n-grammes que Google a rendus disponibles au grand public². Les n-grammes représentent des séquences de n mots dans un corpus de textes. Google a découpé ses pages indexées en ensemble de deux, trois, quatre ou cinq mots, respectivement les deux-grammes, trois-grammes, quatre-grammes et cinq-grammes. À chaque n-gramme est associée une valeur indiquant son nombre d’occurrences dans le corpus. Le corpus de Google utilisé ici prend environ 500 Go d’espace mémoire et date de juin 2013. Tous les n-grammes qui contiennent des nombres ont été pré-filtrés, pour des considérations d’espace disque. Le nom composé *basketball season* est en soit un deux-grammes (ou bigrammes). On pourrait extraire de l’information des bigrammes, mais on se concentre sur les trois, quatre et cinq-grammes qui peuvent contenir un sujet, un verbe et un complément. Ainsi, le trigramme *basketball begins season* est une séquence qui contient un élément prédicatif, le verbe *begins*. Les patrons d’extraction simple fonctionnent avec les n-grammes comme s’il s’agissait de phrases. Les n-grammes sont disponibles rapidement et en grand nombre, et il est facile de rejeter les n-grammes sans intérêt. Butnario & Veale (Butnariu and Veale, 2008) utilisent similairement les n-grammes en cherchant ceux qui commencent par un certain nom suivi d’un verbe et du second nom.

On peut aussi tirer des conclusions de la présence, de l’absence, ou de la fréquence relative d’un groupe de mots cherché.

Le troisième corpus qui pourrait remplacer le moteur de recherche web est Wikipédia. Sans s’approcher de la taille du corpus de Google, Wikipédia a l’avantage d’avoir des phrases bien structurées sur une variété de sujets. Wikipédia est aussi disponible hors-ligne³. Nous

1. <http://jsoup.org/>

2. <http://storage.googleapis.com/books/ngrams/books/datasetv2.html>

3. <https://dumps.wikimedia.org/enwiki/>

utilisons la version datée de mai 2014. L'ensemble des articles en anglais occupe dix gïbiocets en format compressé, et 75 gïbiocets, si rendu disponible sur un serveur. Pour continuer avec l'exemple de *basketball season*, une phrase typique de Wikipédia est *Men's and women's basketball fills the winter season, and spring sports include baseball, fastpitch, and track and field.* tirée de la page de Seattle Lutheran High School. On extrait de cette phrase le verbe *fill*. Nous traitons toutes les combinaisons de pluriel/singulier dans le nom composé. Nous nous basons sur le titre de la page Wikipédia pour éviter d'examiner une page plusieurs fois. Contrairement à Yahoo! et l'approche de Nakov, le nombre de pages pour un nom composé n'est pas limité à 1000 parce que la qualité de Wikipédia est constante, peu importe l'ordre des pages obtenues pour une requête. Une page est jugée pertinente si les deux mots sont situés au plus à 30 mots de distance l'un de l'autre, parce qu'il est préférable d'éliminer le moins de candidats dans ce premier tri. Tout le contenu de la page sera analysé par la suite. Le quatrième et dernier corpus qui sera testé est Faroo⁴, un autre moteur de recherche accessible par requêtes automatisées et gratuit. Son fonctionnement est sensiblement le même que pour Yahoo Boss. Une requête contenant la paire de noms est envoyée au moteur de recherche et le résultat retourné est analysé. On visite les pages web en résultat. Malheureusement, par erreur et contrairement à Yahoo, nous n'avons pas utilisé le résumé de Faroo si la page était inaccessible. Les pages web de Faroo sont traitées de la même manière que celle de Yahoo : le contenu au format HTML est transformé en texte. Pour *basketball season*, la page web <http://dailycaller.com/2013/03/14/grambling-state-basketball-team-completes-winless-season/> contient la phrase *Grambling State basketball team completes winless season* de laquelle il est possible d'extraire le verbe *complete*.

5.2.2 Préparation de phrases en vue de l'analyse

Il y a une série de préparatifs pour obtenir des phrases propres à être analysées à partir de textes tirés des corpus, souvent de qualité variable. Le texte traversera trois étapes : le filtrage des doublons, la segmentation des documents en phrases et le pré-filtrage des phrases sans potentiel.

D'abord, il faut déceler les doublons, quand un texte et sa copie sont traités comme deux textes différents. Il faut s'assurer que le même texte ne soit pas soumis deux fois à l'analyse parce que le poids des éléments prédicatifs qui en sont extraits augmenterait sans justification. Pour les moteurs de recherche Yahoo! et Faroo, l'adresse URL sert à établir l'identité d'une page et à établir que deux pages ont la même source. Finalement, nous utilisons le titre des pages pour repérer les pages Wikipédia qui se répètent. Les n-grammes tout entiers, par leur

4. <http://www.faroo.com/>

petite taille, sont utilisés pour établir l'identité du n-gramme et reconnaître les doublons. De plus, Google a déjà regroupé les n-grammes identiques. Il ne faut qu'être prudent en rassemblant les résultats de requêtes similaires.

Ensuite, il faut séparer les documents en phrases. Chaque corpus est différent. S'il est facile de considérer chaque n-gramme comme une phrase, il faut du temps pour isoler des phrases à partir de pages web ou d'articles de Wikipédia. Les documents seront d'abord divisés en phrases selon les éléments de ponctuation du tableau 5.1. La façon d'indexer Wikipédia force la considération de caractères particuliers, propres aux marques de mise en forme des articles et des info-boîtes. Cette façon de segmenter Wikipédia risque d'écarter de bonnes phrases en les découpant selon un mauvais délimiteur, mais cette procédure évite d'étouffer l'analyseur syntaxique avec de trop longs textes (plus que 50 mots). Une approche textuelle pour segmenter les documents est simple et rapide. Les phrases seront analysées de façon plus poussée à l'étape suivante. Nous conservons le caractère de fin de phrase à la fin de la phrase pour fournir tous les indices possibles pour l'analyse subséquente.

Tableau 5.1 Séparateurs de phrases pour différents corpus

Wikipédia	Faroo	Yahoo !				
.	.	.	\u0000	\u000B	\u1003	\u100C
!	!	!	\u0001	\u000C	\u1004	\u100D
?	?	?	\u0002	\r	\u1005	\u100E
}}	.	.	\u0003	\u000E	\u1006	\u100F
{{	\u0004	\u000F	\u1007	
			\u0005	\u007F	\u1008	
[<	<	\u0006	\u1000	\u1009	
=	>	>	\u0009	\u1001	\u100A	
*		•	\n	\u1002	\u100B	

5.2.3 Expansion par synonymes

L'expansion de synonyme est une étape facultative. L'expérience peut être faite en n'utilisant que les deux noms dans le nom composé. Elle est utilisée pour augmenter le nombre de textes reliés au nom composé recherché. Certains noms composés comme *sob story* ne sont représentés que par quelques textes. La variation de synonyme permet d'accepter les textes de corpus pour *weeping story* et *tear narrative* en plus des textes de *sob story*, non modifiés. Les synonymes affectent la quantité des éléments prédictifs extraits du texte et leur qualité. Leur rôle est d'augmenter la taille des vecteurs d'éléments prédictifs. L'expansion par synonyme n'est acceptable que si on suppose que de remplacer un nom par son synonyme dans le nom

composé n’aura aucune incidence sur la pertinence des éléments prédicatifs ni sur la relation sémantique devinée au final. Pour vérifier cette supposition, nous répéterons l’expérience soit sans synonyme, soit avec des synonymes de différentes sources. Si une expansion est faite, elle utilise soit les liens de Wordnet, soit les synonymes trouvés sur la page de Wordnet ou soit le plus de synonymes possible, sans restriction.

La première approche est celle qui n’utilise que les liens de Wordnet. Wordnet⁵, est « une grande base de données lexicale pour l’anglais. Noms, verbes, adjectifs et adverbes sont regroupés en ensembles de synonymes cognitifs qui expriment chacun un concept différent. » (traduction libre) (Princeton University, 2010). C’est une vaste toile où les noeuds sont des mots et les liens, des relations. Il y a des relations de synonymie, entre *fish* (*for compliments*) et *angle* (*seek indirectly*) par exemple. Il y a des relations d’hyponymie, comme le lien entre *fruit des champs* et *framboises* (en anglais *berry* → *raspberry*). Il y a aussi des relations d’hyperonymie, l’inverse de l’hyponymie, qui relie les mots *framboise* à *fruit des champs* (en anglais *raspberry* → *berry*). L’usage principal de Wordnet est de trouver automatiquement des synonymes. Fournir le mot *dog* permet de trouver 20 synonymes, pour 7 concepts différents, selon qu’il est question de *hot dog*, de *domestic dog*, de *frump*, etc.

Dans la littérature, les utilisations de Wordnet sont nombreuses. Il est utilisé par le système de Girju et son équipe (Girju et al., 2009), qui vont jusqu’à utiliser le même système d’encodage d’information que Wordnet. Le système de Moldovan (Moldovan et al., 2004) utilise les liens d’hyponymie et d’hyperonymie pour clarifier la relation sémantique de ses noms composés. À la tâche 4 de la compétition de SemEval, en 2007⁶, le système avec la meilleure précision, celui de l’équipe de Beamer (Beamer et al., 2007), utilisait Wordnet (Nastase et al., 2013). Selon Nastase (Nastase et al., 2013), Wordnet est une bonne manière d’augmenter le nombre de résultats fournis par un corpus de texte, si en plus de chercher le mot initial, on cherche les synonymes du mot, trouvés automatiquement. En suivant les liens de synonymie à partir des deux noms du nom composé analysé, la méthode de Nakov permet d’augmenter le nombre de noms composés qui expriment la relation sémantique recherchée et donc d’augmenter le nombre de phrases qui rendent explicite cette relation. Cette méthode a l’avantage d’être entièrement automatique et d’être facile à implémenter, ce qui n’est pas le cas des deux prochaines variations. Par contre, le sens dans lequel est employé chacun des noms est inconnu. Par conséquent, plusieurs des synonymes forment des noms composés inusités, qui ne rapportent aucune phrase additionnelle.

Pour les deux dernières approches, nous supposons un moteur de recherche de synonyme par-

5. <http://wordnet.princeton.edu/>

6. <http://nlp.cs.swarthmore.edu/semeval/>

fait. Même s'il est pour l'instant simulé par le jugement d'un humain, apporter des synonymes idéaux dans le processus indiquera s'ils contribuent au succès du processus d'extraction de la relation, et dans quelle mesure. Nous choisissons les synonymes les plus prometteurs de différentes sources, selon l'approche.

Pour la deuxième approche, les synonymes possibles sont choisis à partir du contenu de la page web de Wordnet. Elle inclut les synonymes, mais aussi des définitions, des exemples de phrases et des catégories pour le sens du mot. Nous choisissons les synonymes à partir des mots partout dans la page. Tous les rôles dans la phrase (nom, verbe, adjectif) sont sur la même page. Les synonymes ont été choisis en préservant le plus possible la relation sémantique du nom composé d'origine, uniquement si leur utilisation pouvait améliorer les résultats de l'expérience.

La troisième et dernière approche ne fixe aucune limite sur l'origine des synonymes acceptés par le moteur de recherche idéal. L'accès à toutes les ressources du web, même extérieures à Wordnet, est permis au juge des synonymes, mais le nombre maximal de synonymes par mot est de 10. Le seul mot qui possède autant de synonymes est *morning*, avec *morn*, *forenoon*, *dawn*, *dawning*, *aurora*, *daybreak*, *sunrise*, *dayspring*, *cockcrow* et *sunup*.

5.3 Patrons

Le deuxième module de l'expérience, celui des patrons, inclut l'analyse des phrases, l'application de patrons et l'extraction des éléments prédicatifs. Cette section décrit d'abord l'information gagnée par l'analyse lexicale des phrases, puis décrit les éléments prédicatifs qui seront produits par tout ce processus. Ensuite, elle explore la structure et les capacités d'un patron individuel. Par la suite, nous étayons la variété de caractéristiques pour les patrons de l'expérience. Pour terminer, nous tentons de regrouper les patrons pour extraire le plus d'éléments prédicatifs possible.

5.3.1 Éléments prédicatifs au final

Le résultat des patrons est les éléments prédicatifs : ils sont soit des verbes, soit des verbes jumelés avec la particule qui les suit. Tout comme Nakov, nous excluons les mots auxiliaires typiques des groupes verbaux, comme *will*, *can*, *would*, etc. Par simplicité, et contrainte de temps, nous n'avons pas extrait de prépositions seules ni des conjonctions. Selon les résultats de Nakov, ces éléments contribuent bien moins que les verbes à identifier la relation de Levi. Nous trouvons le lemme des verbes, pour mieux les comparer, à l'exception des formes passives, où nous conservons la forme lemmatisée de l'auxiliaire *be* et la forme fléchie du

participe, afin de distinguer ce cas des formes actives. Nous nous comparerons bien sûr à l'expérience de Nakov qui partage les mêmes conditions.

Au maximum, les éléments prédicatifs sont composés de trois mots : *be*, pour les verbes au passif, le verbe et une particule. Il y a toujours un verbe dans l'élément prédicatif. Nous n'utilisons plus la phrase d'origine une fois l'élément extrait, sauf que nous nous souvenons de sa *direction*. Notre définition de la direction de la phrase est la même que celle de Nakov. Quand les noms du nom composé $N_1 N_2$ apparaissent dans une phrase dans l'ordre $N_1 \dots \text{verbe} \dots N_2$, la direction de la phrase est $N_1 N_2$ ou *avant*. À l'inverse, la phrase $N_2 \dots \text{verbe} \dots N_1$ est de direction $N_2 N_1$ ou *arrière*. Il est essentiel de se souvenir de cette direction puisque la direction est la seule différence entre la relation sémantique **CAUSE**₁ et **CAUSE**₂, par exemple. La direction est aussi utile pour établir une équivalence, si nécessaire, entre un verbe à la forme passive dans une phrase $N_1 N_2$ et un verbe à la forme active dans une phrase $N_2 N_1$.

5.3.2 Analyse syntaxique de surface

Une première étape d'analyse rapide des phrases, caractère par caractère, n'est pas absolument nécessaire, mais permet de gagner du temps en n'invoquant pas l'analyseur syntaxique pour chacune des phrases dans les textes. En général, le corpus nous a fourni un texte avec une ou deux phrases utiles, mais la segmentation en phrase a laissé des phrases qui ne sont pas utilisables. On garde celles qui contiennent les deux noms séparés par au moins un mot. Avec le nom composé *desert rat*, par exemple, la phrase *The rat lives in the desert* est acceptée, mais ni *The desert rat lives*. ni *The rat desert is to the South*. (dans le sens inverse) ne sont acceptées. Une phrase candidate doit de surcroît toujours contenir un des deux mots après avoir rayé de la phrase les apparitions des deux noms juxtaposés, dans un sens ou dans l'autre. Par exemple, la phrase *the honey produced by a honey bee* serait conservée. Ce sera à l'analyse complète, avec patron, de déterminer si la phrase est vraiment utile.

Une fois qu'on a des phrases prometteuses, nous analysons les phrases individuellement en appliquant sur chacune une série d'outils. D'abord, nous passons une phrase dans un segmenteur de phrase par expression régulière. C'est la première étape des pipelines d'analyse de texte typiques. Ensuite, nous identifions les mots avec un segmenteur de mots classique. Suit l'identification de la catégorie grammaticale de chaque mot, autrement dit, sa nature, par un étiqueteur grammatical. En quatrième vient un analyseur morphologique qui fournit les lemmes de tous les mots dans la phrase. Finalement, la phrase est entrée dans un analyseur syntaxique pour identifier les syntagmes (ou groupes fonctionnels) comme les syntagmes nominaux ou les syntagmes verbaux. Il s'agit ici d'analyse syntaxique de surface. Tous les outils

sont fournis par ANNIE, le système d'extraction d'information de Gate⁷, à l'exception de l'analyseur syntaxique *openNlpChunker*⁸, qui provient d'OpenNLP. Nous appliquons presque telle quelle la technique de Nakov à cette étape.

Nous ne savons rien des dépendances entre les groupes fonctionnels. Il faudrait faire une analyse profonde pour découvrir la structure de la phrase en arbre de dépendance. Toutefois, la série d'analyse nous révèle déjà beaucoup. Nous connaissons, grâce à l'analyseur morphologique, le lemme de chacun des mots identifié par le tokéniseur, en plus de son orthographe dans le texte. À titre d'exemple, le lemme du mot *travaux* est *travail* et en anglais, le lemme du mot *made* est *make*. Nous savons aussi la nature du mot, ce qui sera pratique pour distinguer le nom *sobbing* quand il est à la suite d'un verbe et le mot *sobbing* qui est utilisé comme verbe dans la phrase *My cat is sobbing*. Comme dernier atout, grâce à l'analyse syntaxique de surface, nous connaissons l'appartenance de chaque mot à un syntagme, ainsi que s'il est au début du groupe ou non. La nature et le syntagme des mots sont exprimés par les expressions du Penn Treebank Project (Santorini, 1990).

Le nombre grammatical des noms du nom composés n'a plus d'impact à l'étape du filtrage des phrases. Toutes les combinaisons de singulier et de pluriel pour les deux noms des noms composés sont acceptées dans les phrases. La comparaison avec les patrons se fait par les lemmes. La paire de noms devient une paire de lemmes et l'on cherche les lemmes correspondants parmi tous les lemmes de la phrase.

Une fois que la phrase est divisée en mots et que chaque mot est identifié par son lemme, sa catégorie grammaticale et le syntagme auquel il appartient, on est enfin en mesure d'extraire un élément prédicatif de la phrase.

5.3.3 Les patrons et leur expression régulière

Un patron est un motif qui sera détecté dans un texte. Par exemple, le patron **Le chat** ____ appliqué sur le texte **Le chat bleu. Les chats noirs. Le chat dort.** s'appariera avec le groupe de mots **Le chat bleu** et avec **Le chat dort.** On peut extraire des résultats du patron les mots **bleu** et **dort**. La même technique sera appliquée pour trouver nos éléments prédicatifs. Le patron final détectera une portion de phrase avec laquelle il partage des éléments et il isolera des mots d'intérêt. Les patrons peuvent être très généraux, mais on les préfère précis pour assurer la pertinence des éléments prédicatifs extraits. Les patrons ignorent tous les mots qui précèdent ou qui suivent le groupe de mots qu'ils recherchent.

Pour rechercher des groupes de mots dans l'ensemble des phrases du corpus, nous avons

7. <http://gate.ac.uk/>

8. <http://gate.ac.uk/doc/plugins.html#OpenNLP>

utilisé le langage des expressions régulières.

De prime abord, il faut présenter le rôle de certains caractères dans les expressions régulières : Les crochets [et], l'astérisque *, le point . et l'accent circonflexe ^ . Écrire le patron d'un caractère *e* s'appariera avec tous les *e* d'un texte, mais pour trouver toutes les voyelles d'un texte, il faut utiliser les crochets pour écrire le patron [aeiouy]. Ce patron s'apparie avec un seul caractère, qui peut être un *a*, un *e*, un *i*, ... Pour trouver tous les caractères qui ne sont pas des voyelles, le patron [^aeiouy] est construit en ajoutant l'accent circonflexe au début de la liste. Si l'on veut détecter le mot *chat*, l'exclamation *chat!*, le cri *chat!!* et encore *chat!!!!!!*, on utilise l'astérisque * dans le patron *chat!**. L'astérisque (ou étoile de Kleene) permet zéro, une ou plusieurs apparitions du caractère précédent. Pour forcer le caractère à apparaître au moins une fois, le plus + est disponible. Ensuite, le point . remplace n'importe quel caractère unique. Le patron . s'appariera avec tous les caractères du texte, y compris les espaces, les points et les soulignés _. Finalement, si l'on veut obtenir chaque référence de l'expression *Nastase*[1], *Rosario*[2] et *Girju*[?], il faut utiliser la barre oblique inversée. La barre oblique inversée \ évite l'interprétation d'un caractère spécial des expressions régulières et considère le caractère lui-même. Le patron \[.\] combine la barre inversée et le point pour obtenir [1], [2] et [?].

Pour que les annotations comme la nature d'un mot soient accessibles au patron en même temps que le mot lui-même, nous concaténons les quatre caractéristiques de chaque nom comme ceci : *nature:syntagme:lemme:mot*. Si l'on ne connaît ni la nature, ni le syntagme et ni le lemme d'un mot, on peut écrire le mot seul et remplacer les autres annotations par n'importe quelle chaîne de caractères. Pour remplacer n'importe quelle chaîne de caractères, on peut combiner le caractère inconnu et l'astérisque pour dire « aucun ou plusieurs caractères inconnus » : .* . Le patron pour le mot *chats* est .*:.*:.*:chats. On peut faire de même avec les autres catégories. Le patron pour tous les verbes est VB:.*:.*:.*, pour les membres d'un groupe nominal .*:NP:.*:.* et pour capturer *travail* ou *travaux* : .*:.*:travail:.*. Pour contenir les chaînes de caractères inconnues aux caractères qui peuvent apparaître entre les deux points, permettons aux chaînes de caractères inconnues de contenir tous les caractères sauf le deux-points : [^:]* . Il est maintenant possible d'écrire le motif

```
(?: [^:]* : [^:]* NP:word: [^_]* )
```

où les parenthèses isole le motif du reste du patron et où le ?: signifie que le motif entre parenthèses doit être trouvé, mais non extrait de la phrase. Le patron recherche tous les mots qui sont membres d'un groupe nominal (indiqué par une expression se terminant par NP, comme nous le verrons plus loin) et dont le lemme est *word*.

À présent, il est possible d'étendre l'explication au reste de la phrase pour la raison du tiret bas qui est interdit dans la dernière chaîne de caractères inconnue ($[\wedge_]*$). Nous séparons chacun des mots et son groupe de 4 annotations des autres par un tiret bas. Le caractère après la quatrième annotation est un tiret bas parce que la fin du mot est atteinte. La chaîne de caractères inconnue est donc limitée par le tiret bas ($_$). En guise d'exemple, chercher toutes les variations de lemme de *le chat bleu* s'écrit

```
(?:[\wedge:]*:[\wedge:]*:le:[\wedge\_]*)\_(?:[\wedge:]*:[\wedge:]*:chat:[\wedge\_]*)\_
(?:[\wedge:]*:[\wedge:]*:bleu:[\wedge\_]*)
```

Si l'on ignore le nombre de mots à la création d'un motif, il est possible de faire répéter zéro ou plusieurs fois un sous-motif en faisant suivre les parenthèses d'un astérisque : $(?:)*$. Comme nous savons que les mots seront suivis du caractère de séparation, du tiret bas, il est possible de l'insérer dans le sous-motif. Ainsi, pour s'assurer de la présence soit de *le chat*, *le bleu chat* ou *le bleu bleu chat*, il est acceptable d'écrire :

```
(?:[\wedge:]*:[\wedge:]*:le:[\wedge\_]*)\_(?:[\wedge:]*:[\wedge:]*:bleu:[\wedge\_]*\_)*
(?:[\wedge:]*:[\wedge:]*:chat:[\wedge\_]*)
```

L'écriture des patrons pour l'expérience ne nécessite plus que deux subtilités. La première est la possibilité de distinguer les mots qui commencent un nouveau syntagme, comme *Le* et *dort* dans la phrase *Le chat bleu dort*. *Le* commence le groupe nominal *Le chat bleu* et *dort* commence et termine son groupe verbal. Cette distinction est permise grâce à l'analyseur syntaxique, qui appose l'annotation B-NP sur le premier mot des syntagmes et I-NP sur tous les autres. Pour trouver un mot dont le lemme est *word* et qui commence un groupe nominal, on utilise

```
(?:[\wedge:]*:B-NP:word:[\wedge\_]*)
```

Pour trouver le même mot, mais au milieu ou à la fin d'un groupe nominal, on utilise l'annotation I-NP

```
(?:[\wedge:]*:I-NP:word:[\wedge\_]*)
```

Finalement, il est fréquent qu'on soit indifférent à la position d'un mot dans son groupe. Le patron qui correspond à nos besoins est :

```
(?:[\wedge:]*:[\wedge:]*NP:word:[\wedge\_]*)
```

La dernière subtilité concerne l'adaptation des patrons aux différents noms composés recherchés. Notre expérience interprète tous les patrons en remplaçant les mots *word* par le lemme d'un nom du nom composé. Ainsi, avec le nom composé *deserts rats* dont les lemmes sont *desert* et *rat*, le patron simplifié

$(?:[\wedge:]*:[\wedge:]word:[\wedge_]*)_?(?:[\wedge:]*:[\wedge:]word:[\wedge_]*)$

est remplacé d'abord par

$(?:[\wedge:]*:[\wedge:]desert:[\wedge_]*)_?(?:[\wedge:]*:[\wedge:]rat:[\wedge_]*)$

Une copie du patron est ensuite faite avec le remplacement opposé :

$(?:[\wedge:]*:[\wedge:]rat:[\wedge_]*)_?(?:[\wedge:]*:[\wedge:]desert:[\wedge_]*)$

Si le premier patron détecte un élément prédicatif, on note la direction $N_1 N_2$ de la phrase source et pour les éléments détectés par le second patron, la direction notée sera $N_2 N_1$.

Voici finalement un patron complet pour une phrase de la forme sujet-verbe-complément, où l'élément prédicatif extrait est un verbe.

$(?:[\wedge:]*:[\wedge:]NP:word:[\wedge_]*)_?(?:[\wedge:]*:[\wedge:]NP:[\wedge:]*:[\wedge_]*)_*$
 $(?:[\wedge:]*:[\wedge:]VP:[\wedge:]*:[\wedge_]*)_*(VB[\wedge:]*:[\wedge:]VP:[\wedge:]*:[\wedge_]*)_*$
 $(?:[\wedge:]*:[\wedge:]NP:[\wedge:]*:[\wedge_]*)_*(?:[\wedge:]*:[\wedge:]NP:word:[\wedge_]*)$

La première ligne s'assure que le premier nom est dans un groupe nominal et permet à d'autres mots d'être présents entre le nom et le verbe s'ils font partie du même groupe nominal. Sur la deuxième ligne, le premier sous-motif permet aussi à des membres d'un groupe verbal de s'intercaler entre le nom et le verbe. Le verbe lui-même est capturé à la fin de la deuxième ligne, par le sous-motif qui n'est entouré que de parenthèses (sans $?:$). La troisième ligne est le reflet de la première. On permet à des membres d'un groupe nominal de précéder le second nom s'ils font partie du même groupe. La phrase peut contenir d'autres mots après le patron ou avant. Des patrons pour des variations diverses de cette phrase seront présentés dans la prochaine section.

5.3.4 Différents types de patrons utilisés

Devant les multiples formes de phrases rencontrées dans les corpus, il faut nombre de patrons pour arriver à extraire suffisamment d'éléments prédicatifs. Chaque patron est spécialisé pour

une structure de phrase. Tous les patrons s'ajoutent au patron de base *sujet verbe complément*. À l'image 5.2, les cinq structures de phrase supportées sont représentées. Chacun de patrons a accès aux mêmes phrases et fournissent une part des résultats. Les patrons que Nakov a utilisés sont encadrés par un trait plein. Le regroupement de patrons qui a le meilleur rappel, qui capture le plus de verbes possible, est en trait pointillé. Chacune des cinq variétés de patrons, qui s'ajoute au patron simple *sujet-verbe-complément* sera expliquée dans les paragraphes suivants.

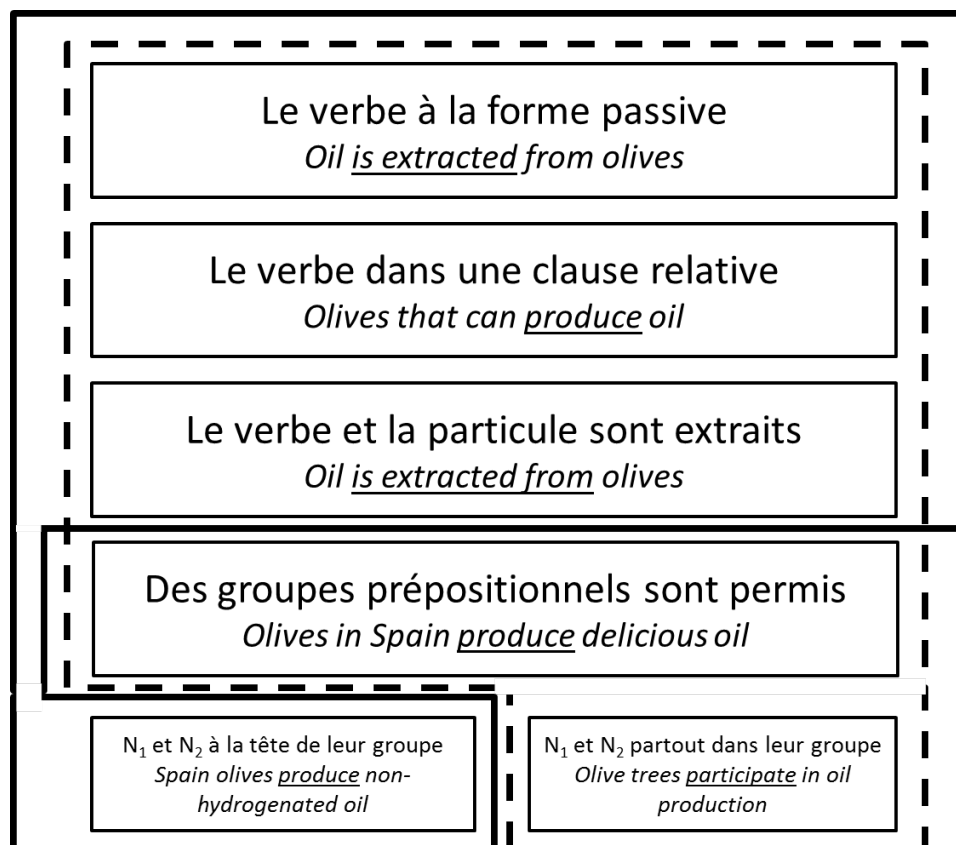


Figure 5.2 Différents ensembles de patrons possibles pour l'extraction d'éléments prédicatifs : *trait plein* : Nakov, *trait pointillé* : meilleur rappel

Premièrement, il est possible qu'un groupe prépositionnel vienne compléter le premier groupe nominal en se positionnant entre le nom et le verbe extrait. Le patron simple n'accepte que le groupe nominal du premier nom et le groupe verbal, alors il élimine la phrase. Pour extraire le verbe de la phrase *A committee from Japan includes many members*, par exemple, il faut un patron qui accepte le groupe *from Japan* dans la phrase. Le groupe qui suit le premier groupe nominal doit commencer par une préposition, dont la catégorie syntaxique est IN et continuer avec un groupe nominal. De plus, le groupe prépositionnel doit être immédiatement suivi du groupe verbal du verbe extrait. Ce type de patron n'est pas nécessaire pour le second

nom, parce que tous les mots qui suivent le second nom sont déjà ignorés.

Deuxièmement, on peut choisir d'accepter les verbes contenus dans une clause relative. Le verbe et son complément sont alors contenus dans une clause relative qui vient décrire le premier nom. La phrase complète continue généralement après la clause relative et compte d'autres verbes. Il est fréquent de trouver des clauses qui débutent par *which* ou par *that*. Une phrase aussi simple que *The committee which includes many members ...* est inutilisée par le patron sujet-verbe-complément. Pour reconnaître la clause relative, on cherche après le premier groupe nominal un seul mot de catégorie WDT, un article en « -wh », ou de catégorie WP, un pronom personnel. Le groupe verbal du verbe à extraire suit immédiatement après.

Troisièmement, comme Nakov, on peut extraire la particule qui suit le verbe en plus du verbe lui-même. La particule est soit de catégorie RP, soit une conjonction de subordination ou une préposition de catégorie IN. Cette stratégie permet d'extraire *rotates off* de la phrase *The committee rotates off $\frac{1}{3}$ of its members*. Cet élément prédicatif est plus spécifique que *rotates*. Si un patron précédent a déjà extrait un verbe simple de la même phrase que le patron qui extrait à la fois verbe et particule, il faut éliminer le premier verbe, pour ne pas le dédoubler. Ce principe s'applique pour tous les patrons.

Quatrièmement, on peut accepter ou non le passif. Cette forme de patron accepte deux mots, *be* suivi par le verbe conjugué au passé. Le patron peut extraire des éléments prédicatifs comptant jusqu'à trois mots : *be*, le verbe au passé et la particule qui l'accompagne. Nous pouvons donc extraire de la phrase *This is a member that is included in many committees* l'élément prédicatif *is included in* en entier. Chaque patron se retrouve donc avec une forme pour les phrases actives et une forme pour les phrases passives. Les éléments à la forme passive sont aussi précieux que les autres puisqu'il est possible de les convertir vers l'actif en inversant la phrase plus tard dans la démarche.

Finalement, il y a un type de patrons qui s'assure de la pertinence de l'élément prédicatif extrait en s'assurant que les noms du nom composé sont à la tête de leur groupe nominal respectif. La position dominante dans un groupe nominal en anglais est à la fin, en dernière position. Si un nom est dans cette position, il est assurément le sujet ou l'objet du verbe de la phrase. Cette restriction permet d'éviter d'extraire de mauvais éléments prédicatifs, mais cause l'élimination de beaucoup de phrases avec un élément prédicatif désirable. Dans un contexte de corpus de relativement petite taille, nous avons tenté l'expérience avec le type de patron opposé, celui qui accepte que les noms recherchés soient à n'importe quelle position dans le syntagme. Plus de bons éléments prédicatifs sont trouvés, mais au coût de mélanger à la liste des éléments qui ne reflètent pas la relation sémantique du nom composé recherchée. Les résultats pourraient être diminués par cette pollution, ou augmentés par

la plus grande quantité d'éléments trouvés. Il y a donc deux types de patrons qui sont mutuellement exclusifs, selon la position du nom dans son syntagme. L'implémentation du type de patron considérant uniquement les têtes de groupe nominal est complexe pour les expressions régulières, particulièrement pour le nom qui termine le patron. Permettre au mot qui suit le nom d'appartenir à tous les types de groupes *sauf* NP n'est pas la fonction première des expressions régulières, d'où la complexité du patron.

Nous avons écrit au total 32 patrons, qui couvrent toutes les variations de clauses relatives, de position dans le groupe nominal, de groupes prépositionnels intercalés, de verbes au passif et d'extraction de verbe et particule. Ils sont tous disponibles à l'annexe D. Il reste cependant beaucoup de formes de phrases qui cachent encore des éléments prédicatifs pertinents. Il serait par exemple possible d'extraire plusieurs verbes d'une seule phrase, telle que *Bees make and store honey* ou d'extraire un verbe à partir de certaines utilisations de gérondif, comme dans les mots *Honey-making bees*. Ces structures de phrases en valent peut-être la peine. Est-ce que tous les types de patron présentés ici en valent la peine ? La prochaine section étudiera la question avec différents regroupements de patrons.

5.3.5 Regroupements de patrons

Dans le contexte d'inaccessibilité de Google, le corpus initialement utilisé par Nakov, il est important que les patrons soient bien choisis, pour utiliser au mieux les nouveaux corpus, plus petits. Il faut distinguer parmi les variations de patrons ceux qui sont utiles. Pour sélectionner les groupes de patrons de l'expérience, trois juges (l'auteur et les directeurs de recherche) se sont penchés sur 497 phrases du corpus de Wikipédia qui concernaient 25 noms composés. Ils ont fait une *pré expérience* pour sélectionner les regroupements de patrons qui ont le plus de potentiel, parmi toutes les combinaisons possibles. Ils ont déterminé si elles contenaient un élément prédicatif désirable ou non. La phrase *These hanging scrolls cover large areas of a wall and usually do not have a border in between* a été choisie puisqu'elle contient l'élément désirable *cover* pour le nom composé *wall scroll*. À l'inverse, la phrase *Rabbi Gedaliah of Semitizi, who went to Jerusalem in 1699, writes that scrolls of the Law were brought to the Wall on occasions of public distress and calamity* doit être ignorée, idéalement, puisque *be brought to* ne révèle pas la relation correcte pour ce nom composé. L'expérience ne se limite pas à vérifier l'extraction des éléments prédicatifs : l'approche doit idéalement sélectionner seulement les éléments prédicatifs utiles pour l'identification. La cohésion entre les juges atteint un kappa de Fleiss (Sørensen, 1948) de 0,45, ce qui est jugé acceptable (Le kappa de Fleiss mesure l'accord entre des juges qui catégorisent les mêmes éléments.) Nous avons déterminé qu'il était désirable d'utiliser une phrase si deux des trois juges se prononçaient

en ce sens.

Les phrases ont été utilisées pour identifier la combinaison de patrons qui accepte un maximum de phrases jugées valides, tout en minimisant le nombre de phrases invalides acceptées. L'approche est binaire, elle ne considère pas la nature de l'élément prédicatif extrait, mais uniquement si les phrases ont été utilisées par les patrons ou non. Les juges ont indiqués que la phrase *The building contains five classrooms for the use of all three grades* est importante pour *classroom use*, mais pas que l'élément extrait doit être *for*. Nous avons calculé la précision (5.1), le rappel (5.2) et la F-mesure (5.3) pour chaque style de patron. Dans les équations, *du* indique le nombre de phrases Désirables Utilisées, *d* indique le nombre de phrases Désirables et *total* indique le nombre total de phrases extraites pour l'ensemble de patrons utilisé.

$$P = \frac{du}{total} \quad (5.1)$$

$$R = \frac{du}{d} \quad (5.2)$$

$$F = 2 \cdot \frac{P \cdot R}{P + R} \quad (5.3)$$

Les résultats sont indiqués au tableau 5.2. L'ajout de certains patrons peut nuire. Leur contribution au score est alors négative. Les résultats, en pourcentage moyen, sont mesurés grâce à la formule 5.4, où *p* est la mesure de performance avec le type de patrons testé et où *t* est la mesure de performance de l'approche qui n'utilise pas le patron supplémentaire (*t* est pour *témoïn*).

$$\frac{p - t}{p} \quad (5.4)$$

Tous les trente-deux regroupements de patrons différents font partie de la pré expérience, même s'ils ne seront pas tous évalués dans l'expérience complète. Les contributions nulles pour les clauses relatives et les phrases au passif ne nous permettent pas de tirer de conclusion satisfaisante sur ces critères. Il n'y avait pas de phrases de ce type dans notre ensemble de phrases choisies au hasard. Permettre les groupes prépositionnels nuit à la précision, mais la F-mesure reste sensiblement la même. Extraire les particules contribue à la moitié du rappel, sans nuire à la précision. C'est un type de patron qui est assurément bénéfique. Sans surprise, ne considérer que les noms qui sont à la tête de leur groupe augmente grandement la précision et nuit au rappel. La F-mesure penche en faveur des patrons de tête de groupe uniquement.

Cette pré-expérience nous a permis de choisir 4 groupes de patrons à tester dans l'expérience principale. Les regroupements sont résumés au tableau 5.3. Nous répliquons bien sûr le groupe

Tableau 5.2 Contributions des types de patrons

Type de patron	Contribution à la précision	Contribution au rappel	Contribution à la F-mesure
groupes prépositionnels permis	-9%	5%	1%
clause relative	0%	0%	0%
particule extraite	5%	52%	41%
forme passive	0%	0%	0%
nom à la tête du groupe	46%	-6%	14%

de patrons de Nakov, PASS/CR/PART/TS, qui extrait les verbes passifs, les particules, les verbes dans les clauses relatives, si les noms sont à tête de leur groupe et qu'il n'y a pas de groupe prépositionnel. Nous formons un regroupement qui maximise la précision. Le regroupement compte un nombre réduit de patrons : PART/TS. Il extrait les particules des noms à la tête de leur groupe et accepte les particules, mais rien d'autre. Un autre regroupement de patrons est formé pour maximiser le rappel : PASS/CR/PART/GP. Il accepte tout : les phrases avec groupe prépositionnel, les verbes dans les clauses relatives, les verbes accompagnés d'une particule, les phrases au passif et même les noms au début de leur groupe fonctionnel. Un quatrième et dernier groupe n'utilisera que les types de patrons qui aident la F-mesure, sans inclure ceux qui n'ont pas d'impact : PART/GP/TS. Il accepte les groupes prépositionnels et les particules, mais oblige les noms à être en position de tête. Les patrons d'extraction de particules sont dans tous les regroupements possibles.

Tableau 5.3 Meilleurs regroupements de patrons selon l'expérience de sélection

Regroupements	passif (PASS)	clause relative (CR)	particule (PART)	groupe prépositionnel (GP)	tête de groupe seulement (TS)
Nakov	✓	✓	✓		✓
Rappel - Témoin	✓	✓	✓	✓	
F-mesure			✓	✓	✓
Précision			✓		✓

5.4 Vecteurs prédictifs de référence

Nous avons comme ensemble de référence 212 noms composés⁹. Pour chacun, un vecteur prédictif décrit la relation entre les deux noms qui forment le nom composé. Les 212 noms

9. Selon l'article (Nakov and Hearst, 2008), les exemples de Levi atteignent le nombre de 214. Or, les listes de vecteurs publiées comptaient 213 noms composés, après n'avoir gardé que les exemples des 12 prédicats

composés sont ceux qui ont été fournis par Levi comme exemples de ses relations, en 1978. Pour chacun, nous avons donc aussi la relation qui leur a été associée par Levi.

En comparant le vecteur prédictif du nom composé inconnu, nous pouvons trouver lequel des 212 noms composés est le plus proche et ainsi attribuer au nom composé inconnu la relation associée à ce vecteur. Pour que cette supposition soit correcte, il est important que les vecteurs prédictifs de référence décrivent exactement la relation de chaque nom composé et que chaque relation soit bien représentée par les exemples. Nous utilisons les mêmes vecteurs que dans l'expérience de Nakov, utilisés à la section 4.4. Il les a rendus disponibles à tous. Deux types de modification sont possibles à partir de ces vecteurs. En premier, les éléments prédictifs, tout comme ceux en provenance du corpus, peuvent être modifiés avant d'atteindre le module de comparaison. En second, il est possible de faire une sélection parmi les éléments retournés par les juges, basée sur l'ordre des réponses dans le formulaire et sur la popularité totale des réponses.

5.4.1 Normalisation et nettoyage des éléments prédictifs

Même si nous n'utilisons pas de juges humains pour produire nos propres vecteurs de référence, nous pouvons tenter d'améliorer l'expérience avec des variations dans laquelle les éléments prédictifs sont préparés différemment. Nous essayons aussi une version de l'expérience dans laquelle les éléments sont normalisés et nettoyés. Les étapes de normalisation et de nettoyage préparent les vecteurs de référence ainsi que le vecteur extrait du corpus, comme il est visible à l'image 5.3. Ces étapes doivent être répétées pour les éléments du corpus de chaque nom composé inconnu, mais ne doivent être exécutées qu'une seule fois pour les vecteurs de référence. Nous avons normalisé et nettoyé les vecteurs de référence à la main, avec le jugement d'un humain, pour préserver la qualité des vecteurs de référence.

En temps normal, les termes des juges proviennent de phrases N_2 N_1 , parce que les phrases à compléter de leur formulaire sont ainsi orientées. Toutes les phrases trouvées sont uniformes, peu importe la direction de la relation sémantique du nom. Il n'est donc pas surprenant de trouver plus de verbes au passif pour $CAUSE_2$ que pour $CAUSE_1$. Nous nous sommes assuré qu'il n'y a que des verbes et des verbes avec particule dans les éléments prédictifs de référence, puisque nous ne répliquons pas l'expérience avec prépositions ou conjonctions.

La normalisation est exécutée avant l'étape du nettoyage. La normalisation se décompose en

de notre expérience. De plus, parmi ces 213, *warrior caste* n'a pas pu être utilisé : il était écrit *warrior castle* sur le formulaire de Mechanical Turk. Nos verbes sont donc légèrement moins nombreux que ceux de Nakov. La liste complète des 212 noms composés tels qu'utilisés est disponible à l'annexe A. Avec l'approche Leave-one-out, où un nom composé à la fois est testé et retiré des vecteurs prédictifs de référence, leur nombre diminue à 211.

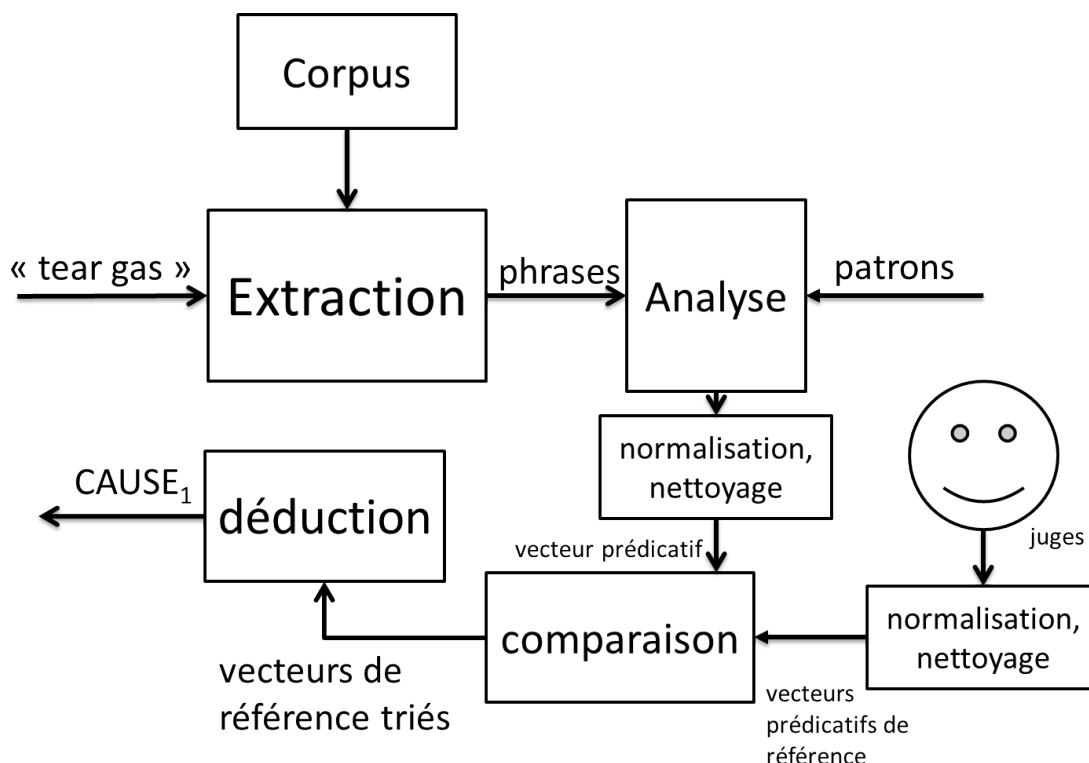


Figure 5.3 Positionnement de la normalisation et du nettoyage des éléments prédicatifs

l'étape de lemmatisation et l'étape d'inversion. La lemmatisation de verbes est automatique pour les verbes de l'engin et est faite manuellement pour les verbes donnés par les juges, pour un meilleur résultat. Nakov choisit de laisser les verbes au passé quand ils sont accompagnés du *be* propre au temps passif, comme *be made of*. À l'opposé, la normalisation change les verbes en leur forme non conjuguée. La normalisation retire aussi toutes les particules, ce qui transformera *be made of* en *be make*. Comme les verbes du corpus doivent être comparés à des verbes de référence au format $N_2 N_1$, nous renversons les verbes passifs trouvés dans une phrase $N_1 N_2$ en verbes actifs de format $N_2 \text{ verbe } N_1$. Inversement, nous renversons les verbes actifs des phrases $N_1 N_2$ en verbes passifs. Après l'inversion, tous les verbes de l'expérience sont interprétés comme $N_2 \text{ verbe } N_1$. Nous supposons que la comparaison entre les éléments du corpus et les éléments de référence peut se faire beaucoup mieux.

En résumé, le tableau 5.4 montre différentes formes du verbe *make* avant et après la normalisation.

L'étape de nettoyage qui suit l'étape de normalisation est plus simple. Il ne s'agit que de corriger les fautes de frappe et autres erreurs remarquées dans les vecteurs de référence, comme passer de *forbodes* à *forebode*. Nous avons aussi supprimé certaines expressions qui

Tableau 5.4 Normalisation

Initial	Normalisé et inversé
$N_1 v N_2$ make	$N_2 v N_1$ be make
$N_1 v N_2$ make of	$N_2 v N_1$ be make
$N_1 v N_2$ be make	$N_2 v N_1$ make
$N_1 v N_2$ be make of	$N_2 v N_1$ make
$N_2 v N_1$ make	$N_2 v N_1$ make
$N_2 v N_1$ make of	$N_2 v N_1$ make
$N_2 v N_1$ be make	$N_2 v N_1$ be make
$N_2 v N_1$ be make of	$N_2 v N_1$ be make

n'étaient pas des termes, comme *be outside of* (sic). À la fin de la normalisation et du nettoyage, le nom composé *queen bee* avait le plus petit vecteur prédicatif, avec 40 termes, et *coriander curry* avait le plus gros, avec 101 termes. Voici un extrait de leur vecteur prédicatif respectif.

Queen bee <be(13), act(5), rule(3), be call(2), reign(2), serve(2), ...>

Coriander curry <contain(18), be make(17), include(10), have(7), taste(5), be flavor(4), come(4), use(3), be(2), be garnish(2), be mix(2), feature(2), smell(2), ...>

En tout, une centaine d'ajustements ont été faits. Tout comme l'étape de normalisation, l'étape de nettoyage est en option et n'est pas effectuée par l'expérience qui réplique la démarche de Nakov telle quelle. Nous utilisons ses vecteurs tels qu'il les a publiés. Il y a donc 3 façons de préparer les verbes : les utiliser sans modifications, uniquement les normaliser ou les normaliser et nettoyer les éléments par la suite.

Selon nous, il est utile de regrouper en une seule instance des éléments similaires à l'aide de la normalisation. La normalisation réunit par exemple *be regarding* et *be regarded* en un seul verbe avec une fréquence de deux, *be regard*. Elle opère à un niveau d'abstraction qui permet de considérer ces deux formes de verbes comme une seule entité. Quand ce verbe est trouvé dans le corpus, il aura un plus grand impact sur la comparaison. La normalisation permet également de comparer correctement les éléments dans le format du formulaire aux éléments trouvés sur internet sans devoir porter attention à l'orientation de leur phrase d'origine. La lemmatisation des verbes au passif permet d'éliminer les erreurs ou les différences de dialectes lors de la mise au passé d'un verbe, comme pour *sew* et ses deux formes : *sewn* et *sewed*.

Enlever les particules est une importante décision, parce que beaucoup de sens est perdu à cette étape. S'il est vrai que cette étape trouve plus d'éléments prédicatifs identiques dans la comparaison entre éléments de corpus et éléments de référence, nous ne pouvons pas être

sûrs que les nouvelles comparaisons sont vraiment souhaitables. Tout comme pour le choix des patrons, il se peut que de sacrifier en partie le sens des verbes extraits de petits corpus se fasse au détriment de la validité de la conclusion finale.

L’approche de Nakov ne semble pas inverser les éléments prédicatifs trouvés. La démarche pour reproduire fidèlement l’approche utilise donc les verbes fournis par Nakov sans normalisation et sans nettoyage additionnel.

Bien que préparés, le nombre de noms composés utilisés comme référence par l’algorithme est relativement petit, ce qui augmente l’importance de bien décrire leur relation, par des éléments prédicatifs minutieusement sélectionnés. La prochaine section explique cette sélection.

5.4.2 Choix des éléments prédicatifs de référence

Le choix des éléments prédicatifs de référence, c’est un questionnement des éléments fournis par les juges humains recrutés par Mechanical Turk. Des vecteurs différents sont assemblés à partir des mêmes réponses. Cette étape n’affecte le parcours des textes, des phrases et des éléments prédicatifs que par l’étape de comparaison, le moment où les vecteurs de référence entrent en jeu. Nous n’avons pas recruté nos propres juges et écrit notre propre questionnaire, mais plutôt repris les vecteurs de référence de Nakov. L’écart entre l’expérience de Nakov et la nôtre est donc limité à cette étape spécifiquement. Il n’y a que trois manières de filtrer les réponses des juges. Les deux approches utilisées par Nakov, utiliser tous les termes ou utiliser uniquement les premiers termes des formulaires, seront essayées. La troisième manière de sélectionner les termes pertinents a pour critère le nombre de fois qu’un terme est mentionné par les juges.

La première version emploie chacun des éléments qui sont donnés par les juges. Cette approche simple maximise la taille des vecteurs de référence. Plus d’éléments dans les vecteurs augmentent les chances qu’un élément trouvé dans le corpus se retrouve aussi dans les vecteurs de référence. Dans certains cas, il est plus facile de trouver un nom composé de référence dont le vecteur est similaire à celui du nom composé inconnu. Utiliser tous les éléments augmente aussi l’écart entre les vecteurs comparés, ce qui permet au nom composé le plus similaire de se distinguer clairement des autres.

La seconde approche testée emploie uniquement le premier des trois termes que chaque juge devait trouver. Cette approche est motivée par l’hypothèse que la première réponse du juge est la plus révélatrice de la relation sémantique. Comme le formulaire exige du juge un minimum de trois termes, la qualité du troisième est mise en doute. Si seuls deux termes décrivent

correctement la relation entre les deux noms du nom composé, comment choisir un troisième terme pertinent ? Par cette logique, la deuxième approche élimine tous les éléments sauf le premier.

La troisième approche s’inspire de l’article de Kim en 2011, (Kim and Nakov, 2011). L’approche de Kim alterne entre l’étape d’extraction automatique des paraphrases de corpus pour des noms composés et l’étape opposée, trouver de nouveaux noms composés qui peuvent être décrits par les paraphrases de la première étape. Kim élimine tous les verbes qui ne sont pas présents au moins cinq fois. Comme notre approche n’est pas automatique et utilise plutôt les jugements combinés de 25 juges, nous réduisons le seuil à deux apparitions. Un seuil plus haut, bien que possible, réduirait considérablement les vecteurs de références. Si un élément n’est mentionné qu’une fois, il peut être exceptionnel et correct, mais il risque surtout de contenir une erreur de frappe ou d’être tout simplement en erreur. Dans tous les cas, il ne représente pas l’interprétation la plus commune du nom composé. C’est selon cette logique que la troisième sélection ne conserve que les éléments prédictifs mentionnés deux fois ou plus.

5.5 Identification de la relation

Cette section décrit le processus de comparaison qui amène au choix final. Le choix final est fait en comparant le vecteur prédictif en provenance du corpus avec les noms composés de référence, après avoir converti chaque fréquence en score TF-IDF. Il y a plusieurs manières de faire cette comparaison, qui sont présentées à la section 5.5.1. Les vecteurs de référence sont comparés soit individuellement, soit regroupés par relation de Levi. La comparaison est faite par coefficient Dice, comme Nakov, ou par distance euclidienne. Il est même possible de baser le choix de relation sur plusieurs noms composés similaires plutôt qu’à un seul.

Pour tenter de prédire laquelle des démarches entraînera les meilleurs résultats, nous avons analysé les vecteurs de référence. La section 5.5.2 présente quelles méthodes de comparaison ont des chances d’amener à de meilleurs résultats et comment elles ont été choisies.

La première étape pour identifier la relation sémantique finale est de remplacer les fréquences de chaque élément en une mesure plus utile, le score TF-IDF. Le score TF-IDF, qui quantifie la rareté de l’élément dans l’ensemble des vecteurs, est utilisé exactement comme il a été décrit dans le chapitre précédent. Un élément qui n’apparaît pas dans le vecteur a un score TF-IDF de 0. Un élément qui apparaît une seule fois a un score qui dépend entièrement de sa rareté dans les vecteurs. Le score TF-IDF pour un élément peut être plus grand que sa fréquence si l’élément est spécifique à un vecteur. La fréquence pure des éléments n’est plus

utilisée pour la suite de l'expérience.

Si le score de comparaison est 0,0 pour tous les vecteurs de référence, aucun nom composé n'est similaire et la relation BE sera choisie parce qu'elle est la plus représentée dans les exemples de Levi.

5.5.1 Extraction de la relation à partir du voisinage

La première des multiples approches pour la comparaison de vecteurs prédicatifs est l'approche de Nakov. La comparaison de vecteur se fait avec la formule du coefficient Dice qui a été présenté précédemment, au chapitre 4. Chacun des vecteurs de référence est comparé au vecteur du nom composé inconnu. Le plus proche voisin est le nom composé de référence qui est associé au vecteur le plus ressemblant. C'est la relation sémantique de ce nom composé qui est choisie comme la plus probable pour le nom composé inconnu.

La deuxième approche emploie la formule de la distance euclidienne au lieu de la formule du coefficient Dice. Cette distance est calculée par la formule euclidienne de la distance classique 5.5, où n est le nombre de dimensions.

$$\sqrt{\sum_{i=1}^n (x_i - y_i)^2} \quad (5.5)$$

La méthode mathématique pour comparer deux vecteurs est différente, mais tout le reste se fait de la même façon. On choisit toujours la relation sémantique du nom composé le plus similaire et tous les exemples de Levi sont comparés un à un.

Une troisième approche, celle des centroïdes, a aussi été mentionnée par Nakov. Si chaque terme représente un axe (orthogonal) d'un système cartésien, et que le score TF-IDF d'un terme est une position sur cet axe, on peut déterminer l'emplacement d'un vecteur prédicatif dans cet espace. Nous choisissons les noms composés de référence qui représentent la même relation de Levi et les positionnons dans l'espace cartésien. Le nuage de points ainsi formé représente le prédicat de Levi. Si d'autres exemples du prédicat étaient trouvés, leur emplacement devrait être à l'intérieur ou à proximité du nuage de points.

Si le nom composé donné au début de la démarche est un exemple du prédicat, il devrait être situé proche du nuage de points. Nous mesurons la distance entre le nom composé inconnu et le centre du nuage pour chaque prédicat de Levi. Mathématiquement, nous utilisons la distance euclidienne comme à la méthode précédente. Le centre du nuage de points est appelé le centroïde. C'est le point qui minimise la somme du carré des distances avec les autres points. Il est calculé par la formule de la moyenne appliquée indépendamment pour chaque

dimension. Une fois les 12 centroïdes trouvés, il suffit de choisir la relation de Levi de celui qui est le plus proche du nom composé en requête.

La quatrième approche tient compte de plusieurs éléments similaires au nom composé inconnu plutôt qu'uniquement le plus proche voisin. La relation de Levi choisie est celle qui est représentée par le plus de noms composés de référence similaires. Si nous prenions en considération les 212 plus proches voisins, notre approche serait équivalente à l'approche de la classe majoritaire et BE serait toujours le résultat en réponse, puisqu'il est représenté par 42 exemples. Si l'approche ne considère qu'un petit nombre de noms composés vraiment similaires, comme trois, cinq et peut-être sept, la relation sémantique est peut-être mieux choisie. Si l'on choisit de considérer les deux noms composés les plus similaires, la relation sémantique du deuxième nom composé ne sera jamais choisie si elle n'est pas aussi celle du plus proche voisin. Nous considérons donc un voisinage de trois, cinq ou sept voisins. La décision est prise par la majorité. Le plus proche voisin tranche dans les cas d'égalité.

Étendre le nombre de noms composés considérés à cinq est risqué si très peu de noms composés sont similaires au nom composé inconnu. C'est pourquoi un nom composé qui ne partage aucun élément avec le nom composé de la requête ($dice = 0$) ne peut pas voter. Avec cinq noms composés, il peut arriver que deux relations soient chacune représentées par deux noms. Dans ce cas, la relation du plus proche voisin l'emporte. Dans le cas où les troisième, quatrième et cinquième voisins ont tous voté pour une relation, c'est celle-là qui est choisie, même si elle est différente de celle du plus proche voisin et du deuxième voisin.

Le voisinage à sept noms composés pousse encore plus loin la confiance dans le nombre de résultats et diminue l'importance accordée à l'ordonnancement des résultats de la comparaison. Tous les autres aspects de l'approche par majorité sont identiques à celle de Nakov : TF-IDF, Dice, comparaison nom composé par nom composé, ...

5.5.2 Analyse des vecteurs pour prédire quel voisinage choisir

Pour tenter de prévoir laquelle des manières de trouver la relation donnera les meilleurs résultats, nous avons fait l'analyse des noms composés de référence. Comme pour l'approche par centroïde, nous avons placé chacun des vecteurs de référence dans l'espace cartésien, pour nous retrouver avec un nuage de point pour chaque relation de Levi. Dans l'évaluation principale de Nakov, Leave-one-out, un point de référence est retiré de l'ensemble de référence et il faut le replacer au bon endroit. Il est donc préférable que chaque nuage de points soit petit et dense. Pour identifier avec confiance la relation sémantique, il faut que chaque nuage de point soit bien défini et distinct de ses voisins. Les zones d'*intersection* entre les nuages de points sont une zone où il est difficile de savoir si un nom composé possède une relation

ou une autre. La forme des nuages et la séparation entre les nuages sont affectées par la manière de choisir la relation sémantique d'un nom composé à partir des noms composés similaires. Il est donc possible de mesurer directement, dans l'espace de points, l'effet de choisir la relation finale par majorité plutôt que par distance du centroïde, par exemple. Il est possible de mesurer la taille des zones d'intersection selon chaque manière de choisir la relation du nom composé inconnu à partir des points de référence.

Les intersections sont mesurées en prenant deux relations à la fois, en comptant le nombre de points d'un nuage qui sont plus proches du nuage de l'autre relation. Dans l'approche par centroïde, un point est considéré dans l'intersection entre deux nuages s'il est plus proche du centroïde étranger que du centroïde principal. L'approche par majorité est simplifiée parce que seuls les points d'un nuage ou de l'autre votent. L'approche par plus proche voisin est tentée en utilisant comme mesure de distance le coefficient Dice et la distance euclidienne dans un second temps.

Nous calculons le ratio entre le nombre de points qui sont classés par erreur dans l'autre relation et le nombre total d'exemples de cette relation. Il est donc avantageux de réduire le plus possible l'intersection entre les deux nuages de points, une zone de confusion. Le pourcentage d'erreurs est calculé pour toutes les 132 combinaisons des 12 relations. Seules les moyennes de chaque mesure, en pourcentage, sont présentées au tableau 5.5. Par exemple, le 2% d'erreur pour l'approche par centroïde avec tous les éléments signifie qu'en moyenne, seulement 2% des noms composés sont jugés plus proche de l'autre prédicat que du prédicat correct.

Tableau 5.5 Intersection moyenne entre les différentes relations de Levi

méthode d'intersection	tous les éléments	premiers éléments	éléments avec plusieurs mentions
centroïde	2%	5%	2%
plus proche, coefficient Dice	3%	5%	4%
plus proche, distance euclidienne	18%	20%	15%
majorité des 3 voisins	20%	23%	19%
majorité des 5 voisins	25%	25%	24%
majorité des 7 voisins	27%	26%	24%
moyenne	16%	17%	15%

Nous avons fait l'analyse pour chacune des trois sélections d'éléments prédictifs de référence de la section 5.4.2 : tous les éléments, les premiers éléments des formulaires et uniquement les éléments mentionnés plusieurs fois. Toutefois, les résultats de cette analyse ne permettent

pas de se prononcer sur la meilleure stratégie. L'approche par majorité entraîne une distinction moins claire entre les catégories que les autres approches, et les zones d'intersection s'agrandissent quand la majorité est calculée avec plus de noms composés, ce qui n'est guère prometteur. Notons l'impact significatif du coefficient de Dice par rapport à la distance euclidienne. Seules les évaluations complètes départageront les deux approches avec le plus de potentiel : l'approche par centroïde et l'approche de Nakov par plus proche voisin. Nous évaluerons l'impact de toutes les versions de cette étape sauf celle de la distance euclidienne dans l'expérience complète, au chapitre 6.

5.6 Résumé des différences avec la méthode de Nakov

Pour conclure ce chapitre, nous présentons la liste des principales différences introduites dans le protocole par rapport à la méthode de Nakov décrite au chapitre précédent.

Le choix des corpus est probablement le point où les différences avec l'approche de Nakov sont les plus prononcées. En effet, l'approche de Nakov strictement équivalente n'est possible qu'en utilisant ses résultats intermédiaires, obtenus sur Google en 2008, quand ce moteur de recherche permettait toujours l'envoi de requêtes construites programmatically dans le choix des corpus. Quatre corpus différents seront essayés : la version anglophone de Wikipédia, les n-grammes web de Google et les deux moteurs de recherche Yahoo! et Faroo.

Deuxièmement, bien que Nakov utilise les synonymes de Wordnet pour augmenter son nombre d'éléments prédictifs, nous allons aussi tester l'approche qui ne permet aucun synonyme et celle qui utilise des synonymes sélectionnés manuellement par un juge.

La troisième section, celle de l'extraction des phrases analysées, est certainement la plus riche en déviations du protocole initial. Une nouvelle sorte de patron est construite pour utiliser les phrases qui comptent des groupes prépositionnels en complément de nom. Seuls les meilleurs ensembles de patrons seront testés, ceux qui favorisent la précision, le rappel ou la F-mesure des résultats lors d'une expérience sur 500 phrases annotées.

Une quatrième différence se retrouve dans la manière de traiter les verbes avant qu'ils soient comparés. Nous essayons d'inverser la forme passive et la forme active d'un verbe pour que tous les verbes proviennent uniformément de phrases où le second nom du nom composé est le sujet et le premier nom est le complément. Nous testons aussi la démarche avec un ensemble d'éléments corrigés.

Cinquièmement, les éléments extraits du corpus sont comparés aux vecteurs prédicats des noms composés de Levi, dont la relation sémantique est connue. Nakov a utilisé deux façons de sélectionner ces éléments prédictifs de référence, une qui emploie tous les éléments et une

qui n'emploie que les éléments mentionnés en premier. Nous réutilisons ces deux sélections, en plus d'une troisième, qui élimine les éléments qui ne sont mentionnés qu'une seule fois.

Sixièmement et dernièrement, nous avons introduit de nouvelles méthodes pour choisir la relation sémantique la plus probable pour le nom composé. Les différentes manières incluent une approche qui regroupe les noms composés de référence par relation sémantique, une approche qui mesure la similarité par distance euclidienne et une approche qui sélectionne la relation de Levi la plus fréquente parmi les noms composés similaires.

De surcroît, de nouvelles manières d'évaluer les résultats de l'expérience seront présentées au chapitre des résultats.

CHAPITRE 6 RÉSULTATS ET ANALYSE

Ce chapitre inclut à la fois nos résultats et une discussion pour, à la lumière des résultats, mieux comprendre les limites de la méthode et pour donner des pistes d'amélioration. Dans la première section se trouvent les résultats pour la simple implémentation de la démarche de Nakov et les résultats d'autres approches. Dans la deuxième section, notre engin est testé par d'autres méthodes d'évaluation. Dans la troisième section, on évalue l'impact des variations faites à différentes étapes de l'approche, qui sont présentées au chapitre précédent. Dans la quatrième et dernière section se trouve la discussion qui contient nos opinions sur les différents résultats et ce qu'il faut en conclure.

6.1 Points de comparaison et résultat

Pour évaluer notre implémentation, nous demandons au système de trouver la relation de Levi pour des noms composés connus et nous comparons ses résultats aux relations connues. Tous nos résultats présentés en pourcentage sont donc des taux de succès.

La méthode plancher, qu'il faut dépasser, n'est pas le hasard, mais la méthode par classe majoritaire. Elle suppose que le nombre d'exemples d'une relation de Levi dans l'ensemble de référence reflète le nombre de noms composés avec cette relation dans la langue anglaise. Elle choisit toujours la relation qui est la plus représentée dans les noms composés de référence, pour augmenter ses chances de deviner la bonne catégorie. La relation BE est toujours choisie. Lors de l'approche Leave-one-out, cette méthode devinera uniquement les 42 noms composés avec la relation BE parmi les 212 noms composés, ce qui lui donne le taux de succès de 20%.

Le plafond suggéré par Nakov pour l'approche Leave-one-out est de 78%. Si nous établissons que l'approche ne peut pas avoir un meilleur résultat que l'accord entre des juges humains effectuant la même tâche, nous atteignons un résultat beaucoup plus bas. Nous avons demandé l'assistance à trois juges pour assigner de nouveaux prédicats de Levi aux 212 noms composés de référence. Le pourcentage d'accord entre les trois n'est que de 54%. Les résultats de cette assignation sont utilisés à la section 6.2.3.

Notre approche témoin, à laquelle est comparée toutes les variations, atteint un taux de succès de 23% en moyenne. L'approche témoin ressemble grandement à l'expérience de Nakov, sauf que le regroupement de patrons employé permet le plus de structures de phrase différentes (meilleur rappel). Avec le corpus de Wikipédia, le résultat est de 23%. Avec les n-grammes de Google, il est de 21%. Il monte à 26% avec Yahoo!. Faroo donne 23% de bons résultats. Ce

résultat, à quelques points au-dessus du plancher de 20%, est très décevant, surtout quand on le compare aux résultats des autres auteurs. Nakov, dans les mêmes circonstances, atteint 43%, soit 20% de plus que notre taux de succès moyen.

Notre expérience la plus similaire à celle de Nakov utilise le moteur de recherche internet Yahoo, ajoute des synonymes parfaits parmi ceux de Wordnet, utilisent les patrons qui obligent le nom à être à la position de tête du groupe nominal et qui ne permettent pas de groupes prépositionnels, n'excluent aucun des verbes des juges de Mechanical Turk et ne considèrent que le plus proche voisin. Cette méthode atteint un résultat de 33%. C'est notre meilleure méthode. L'approche qui s'en rapproche parmi nos variations est celle de Wikipédia, avec un taux de succès de 28%. Elle utilise le premier verbe des formulaires pour peupler les vecteurs de référence. Les éléments de paraphrase sont normalisés.

6.1.1 Comparaison avec d'autres approches

Pour donner des points de comparaison pour l'interprétation des résultats, même si les expériences ne sont pas directement comparables, nous pouvons noter que Rosario et Hearst ont atteint une précision dans le domaine médical de 61%, et ce, pour une classification à 18 catégories. Kim (Kim and Baldwin, 2005) augmente le nombre de catégories à 20 et sa précision atteint 53%. Finalement, les 35 relations de Moldovan (Moldovan et al., 2004) placent sa précision à 43%. Selon Nastase en 2013 (Nastase et al., 2013), les gagnants de SemEval2010 (tâche 9 sur l'interprétation de nom composé par paraphrase), Rink et Harabagiu (Rink and Harabagiu, 2010), ont atteint une précision de 77,9%, améliorant d'un point la précision atteinte par Girju (Girju, 2007) en 2007 de 76% (tâche 4, toujours sur l'interprétation des noms composés). À SemEval 2010¹, selon le site web, un participant atteindrait même une précision de 82%, mais il faut noter que la tâche de SemEval n'a que sept catégories. Si la catégorisation est correcte, un plus grand nombre de relations sémantiques permet d'opérer à un niveau d'abstraction utile, où deux noms composés de même catégorie peuvent être reliés significativement. Pour une classification à 12 catégories, notre seule référence est Nakov lui-même, qui atteint un taux de succès maximum de 50% en utilisant verbes, prépositions et conjonctions comme éléments prédicatifs.

6.1.2 Avec résultats intermédiaires de Nakov sur Google

Nous avons évalué quels résultats peuvent être obtenus à partir des éléments de paraphrase de Nakov tirés de Google en 2008 (Nakov, 2008), s'ils sont donnés directement à notre engin,

1. <http://semeval2.fbk.eu/semeval2.php?location=tasks>

sans devoir interroger de corpus. Ce faisant, nous évitons toutes les étapes pour obtenir des phrases et en extraire des verbes. Les vecteurs prédictifs de Nakov arrivent directement à l'étape de comparaison avec les noms composés de référence. Nous utiliserons le taux de succès de cette expérience pour confirmer les résultats de Nakov. Cette expérience est la plus proche d'une implémentation complète du procédé de Nakov. L'approche fonctionne avec un taux de succès de 42%.

6.2 Différentes méthodes d'évaluation

Nous commencerons par tester notre approche avec d'autres méthodes, pour confirmer nos décevants résultats avant de tenter de différentes variations de protocole pour les améliorer.

6.2.1 Leave-one-out

La première méthode d'évaluation utilise les 212 exemples de Levi. Comme notre ensemble d'évaluation est le même que l'ensemble de référence utilisé par l'algorithme de Nakov, nous devons utiliser l'approche Leave-one-out.

L'approche Leave-one-out consiste à retirer de l'ensemble de référence un nom composé, à imaginer que sa relation est inconnue et à tenter, avec l'engin évalué et avec les 211 noms composés de référence restants, d'identifier sa relation. Si la relation trouvée correspond à la relation que Levi a assignée au nom, c'est un succès. Sinon, c'est un échec. Nous répétons le processus avec tous les autres noms composés de référence pour obtenir au total 212 résultats.

Pour avoir une idée du taux de succès maximum que nous pouvons espérer avec la méthode Leave-one-out, nous l'avons appliquée aux vecteurs de référence obtenus par Mechanical Turk plutôt que d'utiliser les vecteurs obtenus par notre méthode d'extraction. L'hypothèse est que si notre engin fournit une liste aussi complète que des annotateurs humains, il devrait avoir un taux de succès de 100%. Or, en déduisant la relation comme Nakov, nous avons obtenu un résultat de 79%, plutôt que 78%. La différence s'explique probablement par notre ensemble de 212 verbes au lieu de 214. Il se pourrait que l'évaluation Leave-one-out limite le succès maximal. Il serait préférable d'utiliser une évaluation où le meilleur résultat possible est d'au moins 90%, sinon plus.

L'approche Leave-one-out récompense les bonnes approches, mais discrimine aussi selon d'autres critères qui ne devraient pas entrer dans la balance. Tous les cas en échec malgré la qualité de la liste de verbes sont dus à l'utilisation de la relation du nom composé le plus similaire. Tous les cas à la frontière de la définition de la relation sémantique peuvent facilement être classés dans les catégories voisines. Il y a beaucoup de cas frontière avec les

nuages de points dans un espace où chaque verbe est une dimension. Plutôt que de mesurer le succès de l’approche, nous mesurons la cohésion des classes de Levi entre elles, selon leurs exemples. L’espace entre les nuages de points de la bonne et d’une mauvaise relation a un impact direct sur les résultats. Les catégories représentées par seulement quatre candidats sont particulièrement vulnérables. Avoir plus d’exemples de Levi diminuerait l’impact des inconvénients de l’approche Leave-one-out, mais s’il y avait plus d’exemples, nous pourrions aussi avoir un ensemble de tests différents de l’ensemble de référence, et laisser de côté le Leave-one-out. C’est le but de la prochaine méthode d’évaluation : la catégorisation des noms composés de Kim.

6.2.2 Évaluation avec le corpus indépendant de Kim

Nous ferons aussi des tests avec un ensemble de noms composés complètement indépendant de celui de Levi. L’utilisation de nouveaux noms composés, tirés au hasard parmi les noms composés de Kim et Baldwin (Kim and Baldwin, 2005) devrait nous éviter les problèmes liés au Leave-one-out et fournir une évaluation plus standard de notre méthodologie, puisqu’ils sont différents des noms composés de référence. Les noms composés ont généreusement été rendus disponibles sur le site web de Multiword². Nous avons combiné l’ensemble d’entraînement et l’ensemble de tests de Kim pour choisir 417 noms composés parmi les 2 169 que comprend l’ensemble de Kim. Cet ensemble a été séparé en deux fragments de taille similaire. Chacun de ces fragments a été soumis à un juge pour assigner un prédicat de Levi à chaque nom composé. Le nom composé de Kim *agriculture trade* s’est fait assigner les prédicats ABOUT et FROM. De même, le nom composé *state delegation* exprime la relation de Levi HAVE2. La liste complète des assignations est disponible à l’annexe E.

Les noms composés de Kim sont trouvés dans le Wall Street Journal. Beaucoup des termes font référence au domaine de l’économie et de l’investissement et les termes datent de 2005 (plutôt que 1978 pour ceux de Levi). La répartition entre les différents jugements de prédicats est présentée au tableau 6.1. Comme les juges peuvent assigner plusieurs prédicats à chaque nom composé, le total de prédicat est supérieur au total de noms composés. La classe majoritaire est FOR, avec 118 mentions, ce qui amène l’approche plancher, qui identifie toujours la classe majoritaire à obtenir le taux de succès de 20%. Le taux de succès de l’approche plancher avec l’ensemble de Kim ne change presque pas par rapport à l’approche plancher avec l’ensemble de Levi, même si la classe majoritaire n’est plus la même. La classe la moins représentée est CAUSE1, avec cinq noms composés.

2. http://multiword.sourceforge.net/PHITE.php?sitesig=FILES&page=FILES_20_Data_Sets

Tableau 6.1 Répartition des assignations de prédicats de Levi aux noms composés de Kim sélectionnés

Prédicats	nombre d'assignations
cause1	5
cause2	15
make2	23
make1	32
use	32
have1	37
be	50
in	50
have2	61
from	72
about	92
for	118
total	587

6.2.3 Recatégorisation de Levi

Quand Levi a catégorisé des noms composés en 1978, elle n'a jamais affirmé qu'un nom composé ne devait appartenir qu'à une seule catégorie. Elle a même mis en garde les lecteurs contre « tenter de trouver *la* seule interprétation d'un terme » (Levi, 1978). Elle cherchait plutôt à trouver toutes les relations possibles. C'est pourquoi nous avons demandé à 5 juges (l'auteur, les deux directeurs de recherche et des collègues du même domaine) d'assigner une seule catégorie aux 212 noms composés de Levi, sans connaître le jugement de Levi. Les résultats joints à la catégorisation de Levi étaient trop différents pour être utiles, même si le kappa de Fleiss pour les six juges est de 0,52, une valeur acceptable.

Nous avons recommencé le processus, mais avec deux juges seulement (l'auteur et un directeur de recherche) et la possibilité d'assigner de multiples catégories à un même nom composé. Après une première catégorisation, l'ensemble final de catégories a été obtenu par discussion entre les juges. Cet ensemble s'appelle Levi recatégorisé. Beaucoup de noms composés restent identiques, comme *bear country*, dont la relation reste **HAVE**₁ et *gutter language*, qui garde le prédicat **FROM**. Le nom composé *lace handkerchief*, de relation **HAVE**₁ a obtenu une seconde assignation, **MAKE**₂. Quelques noms composés ont été complètement changés. Par exemple, *tire rim* a perdu le prédicat **HAVE**₂ et a acquis les prédicats **HAVE**₁ et **FOR**. L'annexe B regroupe tous les noms composés avec leurs nouvelles relations.

Nous considérerons comme un succès toutes les réponses de l'engin qui sont parmi les catégories plausibles pour le nom composé évalué. Cette évaluation nous permet de vérifier si les

réponses erronées de notre engin sont proches de la bonne réponse. Parce que nous utilisons toujours les mêmes 212 noms composés, nous employons encore l’approche Leave-one-out. Cette méthode d’évaluation rencontre donc les limites typiques de l’approche Leave-one-out. Comme on peut voir au tableau 6.2, la classe majoritaire est toujours BE, avec 46 mentions, ce qui diminue le taux de succès de l’approche plancher à 16%. **MAKE**₁ est la classe la moins représentée, avec six noms composés.

Tableau 6.2 Répartition des nouvelles assignations de prédicats pour les noms composés de Levi

Prédicats	nombre d’assignations
make1	6
cause1	8
cause2	18
for	18
have1	21
have2	21
about	23
in	24
from	28
make2	30
use	44
be	46
total	287

6.2.4 Résultats aux différentes méthodes d’évaluation

Nous avons regroupé les résultats pour les stratégies principales et les différentes méthodes d’évaluation au tableau 6.3. La première approche, celle de Nakov, est la reproduction la plus fidèle possible, avec les mêmes paramètres décrits au chapitre 4. L’expérience témoin est une approche qui est facilement comparable à toutes les variations. Elle n’utilise pas de synonymes, mais utilise les patrons qui maximisent le rappel, pour des vecteurs prédictifs de taille acceptable. La meilleure approche devrait être la combinaison de toutes les variations qui maximisent les résultats à chacune des étapes de la section 6.3.

La meilleure approche n’utilise que les noms du nom composé, sans synonymes. Elle extrait les éléments prédictifs des phrases avec les mêmes patrons que Nakov. Ses vecteurs de référence sont les premiers éléments fournis par les juges de Mechanical Turk, tels quels. Elle inverse les éléments prédictifs du corpus pour avoir la même orientation de phrase. Finalement, elle prend sa décision finale par l’approche des centroïdes, en regroupant les vecteurs de référence

par relation sémantique et en comparant le vecteur du nom composé inconnu à chacun des 12 vecteurs ainsi obtenus.

Trois méthodes d'évaluation serviront à évaluer l'approche : Leave-one-out, les noms composés de Levi recatégorisés et les noms composés de Kim, annotés avec les prédicats de Levi. Utiliser les éléments prédictifs trouvés avec le moteur de recherche Google par Nakov permet d'atteindre 42%.

Tableau 6.3 Résultats pour différentes méthodes d'évaluation

Expérience	Wiki	N-grammes	Yahoo	Faroo	Google
Nakov	22%	24%	33%	21%	42%
Témoin	23%	21%	26%	21%	
Leave-one-out - Meilleure approche	26%	23%	31%	26%	
Levi recatégorisé - Meilleure approche	29%	25%	33%	27%	
Noms composés de Kim, catégorie de Levi - Meilleure approche	19%	19%	21%	18%	
Leave-one-out - Reproduction avec les vecteurs de corpus de Nakov					

Nous analysons les résultats pour les méthodes d'évaluation différentes de celle de Nakov, c'est-à-dire les noms composés de Levi avec de nouveaux prédicats et les noms composés de Kim.

Noms composés de Kim avec prédicats de Levi

Organiser les résultats par prédicats ne révèle aucune corrélation. **FOR** est la classe majoritaire, mais **CAUSE₂** est la classe qui a le plus haut taux de succès moyen. Le prédicat **USE** atteint le plus de documents en moyenne, le plus de phrases en moyenne, mais réussit seulement à deviner 29% des prédicats. Il n'y a aucune corrélation entre les médianes des résultats intermédiaires et le taux de succès.

Le succès ne dépend pas non plus des juges. Le taux de succès moyen pour les résultats des quatre corpus différents est de 19% avec le premier juge et de 17% avec le second juge. Le petit écart de 2% est également visible si l'on ignore les succès dus à l'algorithme de la classe majoritaire. Les résultats intermédiaires ne sont pas affectés significativement non plus.

Les résultats pour les noms composés de Kim partagent la même distribution du nombre de documents et de phrases. Après avoir fait la moyenne des résultats des quatre corpus, nous obtenons une moyenne de 1051 documents par nom composé, beaucoup plus haute que la médiane de 96 documents par nom composé. Le nombre de phrases indique la même

tendance, avec une moyenne de 1856 phrases et une médiane de seulement 412 phrases.

Le taux de succès combiné des quatre corpus est de 18%. Le taux de succès sans classe majoritaire atteint 13%, 5% seulement au-dessus du résultat d'une approche totalement aléatoire. Ces résultats sont égaux ou même inférieurs à ceux de l'évaluation par Leave-one-out. Ils confirment que les faibles taux de succès de notre approche ne sont pas dus à la façon d'évaluer le succès.

Les résultats aux évaluations avec un corpus différent et un classement différents sont prévisibles et amènent aux mêmes conclusions que celles du Leave-one-out.

Analyse des résultats pour la recatégorisation de Levi

Utiliser les noms composés classifiés différemment, avec plusieurs bonnes réponses, plutôt que la catégorisation simple de Levi, améliore les résultats, mais seulement de 3 ou 4%. Il n'est donc pas possible de justifier les erreurs d'identification de prédicat par la possibilité d'attribuer plusieurs prédicats à un nom composé, tous acceptables. Les prédicats fournis par erreur sont sémantiquement éloignés de la réponse attendue.

Ni le nombre de documents, ni le nombre de phrases, ni le nombre d'éléments de paraphrase trouvés ne prédisent le taux de succès des expériences avec les noms composés recatégorisés. Le nombre de verbes qui décrivent les noms composés recatégorisés est corrélé au taux de succès, en excluant les résultats par chance, avec un coefficient de régression linéaire (R^2) de 0,76, mais la même mesure pour les noms composés de Kim descend à 0,35. Une analyse plus détaillée est nécessaire pour atteindre une conclusion sur ce point.

Le prédicat **MAKE**₁ a le meilleur taux de succès, de 58%. À l'opposé, c'est le prédicat **MAKE**₂ qui entraîne le plus d'erreurs en moyenne : 87% des identifications pour **MAKE**₂ sont fausses.

Il y a 19,75 succès (en faisant la moyenne des quatre corpus) attribuables à l'algorithme de la classe majoritaire, parce que le coefficient de Dice n'était pas assez gros. Le nombre de **BE** dans l'ensemble de test est de 184/848, ou 22%. Identifier la relation **BE** quand aucun nom composé de référence n'est proche nous fait gagner 9% de notre succès total. Le nombre de succès total est 61/848, ou 29%. Le nombre de documents en moyenne atteint 726 documents par noms composés, ce qui est très supérieur à la médiane de 86 documents par noms composés. La même situation se produit pour les phrases. Il y a 1282 phrases en moyenne par nom composé, mais la médiane n'est qu'à 344 phrases. Pour chaque nom composé connu du corpus et très bien documenté, il y a donc beaucoup de noms composés introuvables. Il n'y a que 61% des noms composés pour lesquels un nom composé de référence avec un élément prédictif similaire est trouvé. Pour tous les autres, la comparaison de vecteur ne trouve aucun nom

composé similaire.

Les mauvaises attributions de la classe majoritaire causent 623 erreurs, sans surprise. Le prédicat **USE** est la deuxième cause du plus grand nombre d'erreurs. Le programme prédit 158 fois une relation de type **USE** pour un nom composé avec une autre relation.

6.3 Résultats pour les variations du protocole

Cette section rassemble les évaluations de toutes les variations du protocole principal. Chaque section présente pour une étape de la démarche les différentes versions possibles, les résultats et notre jugement devant ces résultats. Tous les taux de succès sont présentés en pourcentage et arrondis au chiffre le plus près. L'évaluation choisie est celle de Leave-one-out, malgré ses failles, par contrainte de temps. L'expérience témoin est présente à chaque tableau. Elle ressemble à l'approche de Nakov, mais simplifiée, puisqu'elle n'utilise pas de synonymes et qu'elle utilise tous les patrons pour maximiser le nombre de verbes récoltés. Tous les vecteurs de références sont utilisés. L'approche du plus proche voisin et du coefficient Dice est employée pour déduire la relation sémantique.

Nous utilisons la classe majoritaire **BE** quand aucun nom composé de référence similaire n'est trouvé. Nous choisissons de présenter les résultats incluant les succès dus à la chance. Ne pas inclure ses succès favoriserait certains procédés et en défavoriserait d'autres. Nous préférons présenter le meilleur taux de succès possible pour que les améliorations soient clairement visibles et que notre approche soit comparée à son plein potentiel aux autres protocoles. Comme l'utilisation de la classe majoritaire est l'algorithme de secours pour toutes les expériences présentées ici, nous pouvons espérer que l'impact des variations individuelles sera tout de même visible.

6.3.1 Variation du corpus

Comme nous avons testé la méthode de Nakov et toutes les autres variations avec les quatre corpus, il nous a suffi de rassembler le taux de succès maximum, moyen et minimum pour chaque corpus dans le tableau 6.4. L'étoile marque une valeur extrême dans la rangée. Toutes les expériences utilisant l'évaluation par Leave-one-out sont incluses dans ces statistiques : l'approche témoin, comptée une seule fois, toutes les manières d'aborder les quatre variations, la meilleure approche et l'approche la plus fidèle à celle de Nakov.

Le taux de succès maximum est de 33%. Il est atteint deux fois. Le premier par l'approche initiale de Nakov avec Yahoo! et le deuxième par l'approche de Nakov, mais sans synonymes, lors du test des patrons. Les patrons semblent avoir une plus grande influence que les syno-

nymes. Le deuxième meilleur taux de succès, 28%, est obtenu avec Wikipédia, en prenant les premiers verbes, normalisés, ou en utilisant l'approche par centroïde.

Notre pire taux de résultat est atteint avec Yahoo!, par centroïde ou avec l'approche utilisant le plus de synonymes.

Wikipédia est notre meilleur corpus en moyenne. Il représente notre meilleure solution après Yahoo!, même si le résultat de 25% est beaucoup plus bas que ce qui est atteint avec les éléments prédicatifs de Google (42%).

Le corpus des n-grammes ne se distingue pas et présente un taux de succès maximum de 25% avec les premiers verbes de Mechanical Turk, nettoyés et normalisés, bien que ce résultat ne soit pas significativement différent des autres variations. Le corpus de Faroo a la pire moyenne, à 20%, en dessous de la moyenne de Wikipédia. Elle atteint exactement l'algorithme de classe majoritaire, notre plancher des résultats (à 20%). Bien qu'il soit le plus rapide à appliquer (une heure), Faroo ne fournit pas assez de documents pour notre application.

Tableau 6.4 Statistiques de performance des différents corpus

Mesure	Wiki	N-grammes	Yahoo	Faroo
maximum	28%	25%	33%*	26%
moyenne	25%*	22%	24%	20%
minimum	21%	20%	17%*	18%

Les performances atteintes par Yahoo! par rapport aux autres corpus appuient l'opinion de Nakov que « les plus petits corpus, même mieux analysés linguistiquement, ne peuvent pas être un remplacement du web, plus grand de plusieurs ordres de magnitude » (Nakov and Hearst, 2013). Toutefois, utiliser Google n'est plus possible et utiliser Yahoo! est coûteux, autant financièrement que par le nombre de requêtes, et le temps qu'il faut pour produire une réponse, d'un à cinq jours pour l'approche par synonyme.

Il semble clair qu'il faut traiter les corpus de différentes façons. L'approche par centroïde améliore grandement la performance de Wikipédia, mais fait dégringoler celle de Yahoo à 17%, son pire résultat. Une approche combinée des corpus de plus petite taille a le potentiel d'atteindre la performance obtenue par Yahoo.

6.3.2 Approche par synonyme

L'approche par synonyme était présente dans la méthode de Nakov pour générer des requêtes supplémentaires au corpus et en tirer plus de phrases pouvant contenir un élément prédicatif pertinent. Pour les synonymes, nous supposons un moteur de recherche de synonyme idéal.

Notre expérience témoin n'utilise pas de synonymes parce que nous n'utilisons pas comme Nakov la façade logicielle de Wordnet pour trouver des synonymes. Un jugement humain a été posé sur les liens de la version internet de Wordnet 3.1³, accédée en été 2014, pour constituer un ensemble de synonymes Wordnet qui ont été sélectionnés pour conserver la relation sémantique avec le second membre du nom composé, connu par le juge humain. Le juge humain applique les mêmes critères de sélection pour constituer un ensemble de synonymes rassemblé à partir de tous les mots de la page internet de Wordnet 3.1 et un ensemble de synonymes rassemblés sans restriction, de source possiblement extérieure à Wordnet.

Les taux de succès pour chaque synonyme et chaque corpus sont regroupés au tableau 6.5. Toutes les variations avec synonyme se sont avérées moins performantes en moyenne que l'approche témoin qui n'en utilise aucun. La plus grande chute a été subie par le corpus Yahoo ! qui est passé du meilleur taux de succès au pire, à 17%, soit trois points sous l'approche plancher. Seuls les n-grammes voient leur succès augmenter légèrement. Tous les résultats sont très proches les uns les autres. Il est clair que toute utilisation des synonymes avec l'approche de Nakov devra être faite différemment. Il nous faut conclure que les résultats de quelques points de plus pour Wikipédia et les n-grammes avec les synonymes de sources variées ne justifient pas de laisser cette étape dans le protocole.

Tableau 6.5 Taux de succès pour différents ensembles de synonymes en requête au corpus

Patron	Wiki	N-grammes	Yahoo	Faroo	moyenne
Aucun synonyme - témoin	23%	21%	26%	21%	23%
Liens Wordnet	21%	23%	22%	21%	22%
Page de Wordnet	24%	24%	19%	20%	22%
Tous les synonymes	23%	24%	17%	18%	21%

En plus de ne pas augmenter significativement le taux de succès, l'approche par synonyme demande environ cinq fois plus de temps. Il a fallu cinq jours pour évaluer le résultat avec tous les synonymes et le grand corpus de Yahoo ! Que s'est-il passé pour expliquer cette chute du taux de succès ? En nous concentrant sur les différences entre ces deux expériences, nous constatons que le nombre de verbes distincts passe de 70 sans synonymes à 185, en moyenne. Les synonymes ont donc bien un impact. La somme des scores TF-IDF, en moyenne pour les 212 noms composés passe de 591 à 2274, parce que plus de verbes contribuent au score. La moyenne et la médiane pour le coefficient Dice du nom composé le plus similaire diminuent légèrement avec l'approche par synonyme, probablement parce que le nombre de verbes trouvés, un des termes du dénominateur, augmente beaucoup plus que le nombre de verbes communs aux deux listes, au numérateur.

3. <http://wordnetweb.princeton.edu/perl/webwn> ?

En regardant uniquement les noms composés qui ne sont plus devinés avec les synonymes, nous constatons le même état des choses. Presque dix fois plus de documents ont été retirés du corpus en moyenne. Le nombre de verbes distincts est bien plus grand et, par conséquent, la somme des scores TF-IDF aussi. Le nombre de verbes communs avec le plus proche voisin de la bonne relation augmente avec l'approche par synonymes. Toutefois, le nombre de verbes communs avec d'autres vecteurs de référence, qui ne partagent pas de relation sémantique, augmente autant et parfois, plus.

L'exemple d'*automobile plant* illustre bien ce fait. Sans synonymes, son vecteur prédicatif compte 30 verbes distincts, dont seulement *include(13)* et *feature(6)* contribuent à la comparaison avec le nom composé le plus proche qui s'avère partager la même relation, *basketball season*.

<include (13), increase(12), feature (6), be(6), build(6), applaud(3), begin(3), establish(3), open(3), sell(3), work(3), launch(2), leave(2), manufacture(2), produce(2), s(2), abandon(1), ask(1), be renamed as(1), breed(1), chronicle of(1), commence(1), comprise(1), cooperate with(1), interest in(1), irve(1), rename into(1), seal(1), spark(1), use(1)>

La relation est bien identifiée. Utiliser les synonymes de toutes origines pour le même nom composé augmente le vecteur prédicatif à 255 verbes distincts, mais la relation devinée n'est plus correcte. Pour *automobile*, les synonymes sont *auto*, *car* et *motorcar*. Pour *plant*, *works* et *buildings* sont des synonymes acceptables. Voici le vecteur sans les éléments uniques :

<be (34), build (33), lead (26), work (24), open (21), s (20), include (16), announce (16), manufacture (16), produce (14), assemble (12), increase (12), make (10), need (9), race (9), see (8), set (8), view (8), feature (7), beij (6), buy (6), get (6), price (6), require (6), sell (6), cover (5), fit (5), sap (5), use (5), begin (4), launch (4), mirror (4), run (4), water (4), wish (4), have (3), applaud (3), ask (3), be assembled (3), be in (3), call (3), compare (3), do (3), drum (3), establish (3), find (3), inspect (3), leave (3), pine (3), relate (3), set up (3), start (3), abandon (2), add (2), air (2), allege (2), allow (2), bank (2), be damaged in (2), be made at (2), be making (2), block (2), bolster (2), browse (2), clean (2), discuss (2), do at (2), drive (2), edit (2), enable (2), endanger (2), explore (2), flash (2), groundbreak (2), heal (2), hold (2), inform (2), install (2), interest in (2), invest in (2), line up (2), live (2), log (2), maintain (2), make at (2), motorize (2), mount (2), pedal (2), provide (2), relate from (2), repair (2), roll out (2), seal (2), service (2), style (2), supply (2), test (2), top (2), train (2), use for (2), use in (2), visit (2), want

(2), win (2), ... >

Pour le vecteur avec synonymes, cinq éléments distincts ont été utiles à la comparaison.

<include(16), feature(7), have(3), contain(1), encompass(1)>

Cette fois, le nom composé le plus similaire est *honey bee*, qui est une erreur, selon la classification de Levi.

Les vecteurs de référence sont trop petits par rapport aux listes de verbes recueillis. Par exemple, *steam iron* compte 2 verbes *run on* qui s'associent à 13 verbes *run on* tirés du corpus. L'élément le plus fréquent dans le vecteur de *steam iron* est *use*, avec 12 apparitions. Il est comparé à 80 apparitions du verbe dans le corpus. Comme l'opération Dice ne considère que le minimum entre 2 et 13 pour *run on* et entre 12 et 80 pour *use*, le score reste petit.

Un score légèrement plus haut est atteint quand *steam iron* est comparé avec le nom composé de référence d'*automobile plant*, qui représente un mauvais prédicat. Le verbe le plus important de ce dernier, *make*, avec 17 apparitions, se compare tout aussi bien avec les 27 apparitions de *make* dans le vecteur de corpus, où 10 éléments prédicats distincts sont plus nombreux encore que *make*. Le minimum est toujours 17, et la plus grande part des éléments prédicatifs gagnés par les synonymes ne contribuent pas. C'est ce qui explique que l'approche par synonyme avec Yahoo! cause la perte de 28 noms composés et le gain de seulement 11 noms.

Il faudrait soit changer la formule Dice pour une formule qui multiplierait les scores TF-IDF normalisés plutôt que d'utiliser le score minimum ou encore bonifier les listes de référence pour être de taille comparable aux nombreux verbes des synonymes.

6.3.3 Équipe de patrons

Varier la façon de sélectionner les éléments prédicatifs affectera sûrement les résultats. Les phrases passives doivent-elles être utilisées ? Les éléments du nom composé doivent-ils être obligatoirement en position dominante ? Pouvons-nous accepter que des groupes prépositionnels soient placés entre le groupe nominal et le groupe verbal dans la phrase ? Acceptons-nous les verbes contenus dans des clauses relatives ? Finalement, est-ce que les particules seules peuvent être considérées comme élément prédicatif, en plus des verbes et des verbes liés à une particule ? Après analyse dans le chapitre 5, nous avons sélectionné quatre façons de répondre à ces cinq questions. Le regroupement de patrons de Nakov reproduit les patrons employés par Nakov. Deux des regroupements favorisent soit la précision, soit la F-mesure à l'expérience de sélection. Le regroupement qui favorise le rappel, et qui fournit le plus de verbes possible, est celui utilisé par l'expérience témoin. La composition détaillée des regroupements

peut être trouvée au tableau 5.3. Les patrons complets sont fournis en annexe.

Tableau 6.6 Taux de succès pour différents patrons

Patrons	Wiki	N-grammes	Yahoo	Faroo	moyenne
Nakov	25%	21%	33%	21%	25%
Précision	23%	22%	27%	20%	23%
Rappel - témoin	23%	21%	26%	21%	23%
F-mesure	25%	22%	25%	20%	23%

L'impact du choix de l'équipe de patrons sur les résultats est plus faible que prévu. Pour Wikipédia, les n-grammes et Faroo, aucune différence significative n'est notée. Les corpus semblent avoir plus d'influence sur les résultats, et Yahoo! est systématiquement meilleur. Seul Wikipédia avec les patrons favorisant la F-mesure atteint le plus faible des résultats de Yahoo! Le meilleur résultat est Yahoo!, le corpus qui ressemble le plus à Google, avec les patrons que Nakov a utilisés. Toute tentative d'amélioration de cette étape par rapport à l'expérience reproduite s'est soldée pour un échec.

Le nombre de phrases fournies à l'analyse et le nombre de verbes que les patrons d'extraction fournissent sont regroupés par corpus au tableau 6.7. Les étoiles indiquent une valeur maximale.

Il est fort possible que les différences entre les regroupements de patrons s'agrandissent avec le nombre de phrases fournies. Yahoo fournit plus de phrases, et Wikipédia semble mieux utiliser ses phrases, puisque plus de verbes sont extraits par Wikipédia que par Yahoo. Ce n'est qu'un cas, mais *tire rim* illustre bien nos propos. Yahoo trouve 4135 phrases pour ce nom composé, soit beaucoup plus que Wikipédia, avec 305 phrases. Pourtant, le vecteur prédicatif de Wikipédia compte 22 verbes (ratio : 5%), comparé à celui de Yahoo, qui compte 56 verbes (ratio : 1%). Le vecteur de Wikipédia pour *tire rim*, sans les éléments uniques, est :

<come off (3), bolt through (2), stay on (2), fit (2), be mounted on (2), ... >

Celui de Yahoo est :

<fit (12), mount on (9), have (7), fit on (6), remove (5), match (4), fall from (4), be dropped from (4), change (3), cause (3), be (3), wrap with (2), use (2), stay on (2), mount onto (2), meet (2), hold (2), go on (2), be on (2), be mounted upon (2), be mounted on (2), ... >

Yahoo a une très haute médiane pour les verbes, ce que nous interprétons comme plus de verbes trouvés pour les noms composés rares. Cette caractéristique pourrait expliquer l'écart de 8% entre Yahoo! et Wikipédia.

Les patrons de Nakov acceptent les verbes passifs et les verbes dans une clause relative. L'expérience qui les utilise est supérieure à celle qui utilise les patrons de précision par un écart de 16 noms composés supplémentaires, où la relation est identifiée par la variation avec les patrons de Nakov, mais pas par les patrons de précision. Dans tous ces cas et aussi dans les 2 cas inverses, les patrons de Nakov extraient plus de verbes en moyenne. L'analyse comparative des vecteurs de verbes révèle que les patrons pour le passif fournissent 16 verbes utiles, et ceux pour les clauses relatives, 15 verbes. Dix noms composés sont trouvés grâce aux patrons passifs et six grâce aux clauses relatives.

Dans l'expérience présentée à la section 5.3.5, les phrases annotées ne contenaient pas suffisamment de phrases pour se prononcer sur l'utilité de considérer les phrases passives et les clauses relatives. Nous savons maintenant qu'il est utile d'extraire un verbe de ces types de phrase pour Yahoo! Même si la moyenne des performances reste très faible, peu importe le corpus, et que l'écart en faveur des patrons de Nakov ne soit pas significatif, l'inclusion des phrases passives et des clauses relatives ne nuisent pas.

Dernièrement, nous notons que permettre des groupes prépositionnels et des noms ailleurs qu'en position dominante n'a pas donné l'amélioration escomptée, malgré le contexte d'attrition de verbe. Ils ne nuisent pas, mais ne contribuent pas non plus. Leur influence sera véritablement testée quand le taux de succès s'éloignera du plancher.

Tableau 6.7 Quantité de documents et de phrases pour les différents corpus avec les patrons de Nakov

corpus	phrases candidates			verbes		
	somme	moyenne	médiane	somme	moyenne	médiane
Wikipédia	265594	1253	269	3916*	18	5
n-grammes	121	1	1	1170	6	1
Faroo	17455	82	32	290	1	0
Yahoo!	397638*	1876*	1616*	3577	17	11*

6.3.4 Variation dans la sélection et le traitement des éléments prédicatifs de référence

Pour améliorer les résultats, il est possible de choisir différemment le prétraitement des verbes de référence de Nakov. Nous pouvons normaliser tous les éléments prédicatifs, de référence comme du corpus, et utiliser un ensemble nettoyé par rapport à l'original. Avec la méthode sans normalisation de l'expérience témoin, il y a trois traitements différents possibles. Nous pouvons aussi, indépendamment du traitement, choisir de sélectionner tous les éléments prédicatifs de référence, ne prendre que les premiers éléments des formulaires ou encore ne

prendre que les éléments récurrents. Comme il y a trois choix de traitement et trois choix de sélection de verbes, neuf expériences différentes seront faites.

Les résultats de ces expériences sont contenus dans le tableau 6.8.

En moyenne, toutes les sélections de verbes ont des résultats décevants à 22% ou 23%. Les moyennes de chaque traitement sont peu révélatrices parce que chaque traitement a la moyenne de 22%. Nous ne pouvons tirer aucune conclusion des moyennes pour chaque expérience individuelle. Si l'on ne considère que les maximums, prendre les premiers verbes est utile pour Wikipédia et les n-grammes, mais il semble préférable pour les corpus web de Yahoo! et de Faroo d'avoir le plus de verbes possible. Les résultats sont partagés pour le traitement des verbes. La normalisation seule fonctionne bien pour Wikipédia, mais n'aide aucun autre corpus. Le nettoyage n'a jamais aidé. Les résultats pour Faroo sont en dessous du plancher de 20% et ne sont pas considérés.

Tableau 6.8 Taux de succès pour différentes sélections des éléments prédicatifs

patron	Wikipédia	Ngram	Yahoo	Faroo	Moy
premier verbe	21%	23%	23%	21%	22%
premier verbe normalisé	28%	24%	21%	18%	23%
premier verbe, nettoyé, normalisé	26%	25%	22%	18%	23%
tous les verbes - témoin	23%	21%	26%	21%	23%
tous les verbes normalisés	27%	20%	22%	18%	22%
tous les verbes nettoyés et normalisés	26%	20%	22%	19%	22%
verbes >1	23%	22%	24%	19%	22%
verbes >1 normalisés	25%	22%	20%	19%	22%
verbes >1 nettoyés et normalisés	24%	21%	22%	18%	21%

Comme il n'y a pas d'amélioration apportée par les différentes sélections de verbes de référence, nous concentrerons nos efforts sur la façon par laquelle la normalisation affecte les résultats. Nous nous sommes penchés sur les gains et pertes de l'expérience avec le plus grand résultat, celle avec Wikipédia et avec les premiers verbes normalisés, par rapport à l'expérience sans normalisation. Nous avons omis de considérer, lors de la normalisation, que le verbe *be* dans une phrase avant l'inversion de cette phrase deviendra *be be*, puisqu'un *be* marqueur de passif a été ajouté au lemme *be* du verbe principal. Aucun de ces verbes ne peut être retrouvé dans les exemples. Il aurait fallu faire un cas d'exception pour ce verbe ou encore comparer les verbes d'une autre façon que caractère par caractère, en tenant compte de la direction $N_1 N_2$ ou $N_2 N_1$ de la phrase, du passif et des lemmes dans une structure plus complexe qu'une chaîne de caractères. Cette erreur explique beaucoup de pertes, et explique pourquoi 12 des 19 pertes sont des noms de catégorie BE. Nous regrettons de ne pas avoir pu

la corriger par manque de temps. L'élision du *have* et de *be* en *'s* entraîne sûrement quelques erreurs. Les gains ont été faits malgré la perte du *be* dans de multiples catégories. Plusieurs exemples observés confirment l'hypothèse derrière la normalisation : l'inversion et la lemmatisation sans particules sont responsable des meilleures réponses, même si les scores TF-IDF sont légèrement diminués par la perte de particule, qui rend les verbes moins rares dans le corpus.

En général, il suffit d'un verbe pour faire la différence. Pour qu'une si petite différence de coefficient propulse les bonnes réponses aussi loin en arrière, il faut que tous les noms composés de référence soient en peloton. Pour les noms perdus, la différence moyenne entre le score de similarité pour la bonne et la mauvaise réponse est de 0,05, alors que le coefficient peut monter jusqu'à 0,42. Les petites différences à la liste de verbes ont propulsé le nom composé réponse sans normalisation à la vingt-quatrième place en moyenne parmi les noms composés de référence, avec la normalisation. Nous croyons par conséquent qu'il nous manque des éléments prédicatifs dans nos vecteurs de référence pour prendre une décision stable et assurée.

Pour l'expérience avec les meilleurs paramètres, nous choisissons de prendre les premiers éléments prédicatifs, normalisés. Ils fournissent le taux de succès maximal avec Wikipédia, mais uniquement avec ce corpus. Aucune version du protocole ne s'est vraiment distinguée à cette étape de l'algorithme.

6.3.5 Méthode d'identification de la relation sémantique

Après avoir augmenté le nombre d'éléments prédicatifs par les synonymes et avoir essayé différents vecteurs prédicatifs de référence, la manière de faire la comparaison finale doit être remise en question. Notre témoin utilise l'approche de plus proche voisin employée par Nakov. Nous essayons d'exploiter les relations sémantiques du trois, cinq et même sept voisins, en trouvant la relation majoritaire, après avoir classé les noms composés de référence en ordre décroissant de similarité. Nous changeons aussi la manière de choisir la relation sémantique en regroupant d'abord les noms composés par relation pour combiner leurs éléments prédicatifs avant la comparaison. Les 12 vecteurs prédicatifs qui en résultent sont placés à la position du centroïde par rapport aux noms composés individuels pour leur relation. La relation du plus similaire est choisie.

Le tableau 6.9 présente les résultats des expériences avec différentes méthodes pour l'identification du prédicat.

La différence entre les moyennes pour les différentes méthodes d'extraction n'est pas suffi-

Tableau 6.9 Taux de succès pour différentes méthodes d'identification de la relation sémantique

Méthode d'extraction du voisinage	Wiki	N-grammes	Yahoo	Faroo	moyenne
Plus proche voisin - témoin	23%	21%	26%	21%	23%
Majorité à 3	26%	22%	25%	18%	23%
Majorité à 5	25%	23%	23%	20%	23%
Majorité à 7	24%	23%	25%	21%	23%
Centroïde	28%	21%	17%	18%	21%

sante pour se prononcer en faveur d'une méthode ou d'une autre. Nous notons cependant que l'approche par plus proche voisin semble particulièrement appropriée pour Yahoo! et que l'approche par centroïde permet d'obtenir le meilleur taux de succès du corpus de Wikipédia. Tout comme l'expérience de variation des patrons d'extraction, il nous faut attendre de meilleurs résultats pour bien distinguer la meilleure méthode d'identification. Comme l'approche par centroïde a fourni le meilleur résultat, c'est celle-là qui est utilisée par l'approche avec les meilleurs paramètres à la section 6.2.4.

6.4 Analyses supplémentaires

Répéter les expériences avec variations nous donne l'occasion de comparer les paramètres, mais nous fournit aussi beaucoup de résultats intermédiaires. Nous utilisons ces indices pour mieux comprendre le fonctionnement de l'algorithme à différentes étapes d'achèvement. Cela nous permettra de tirer des conclusions de la différence entre nos résultats et ceux de Nakov. Nous effectuerons l'analyse des erreurs de l'analyse des phrases, de l'influence du vecteur en provenance du corpus et de l'influence des noms composés du prédicat BE.

6.4.1 Erreurs dans l'extraction des éléments prédicatifs

Il y a des erreurs commises par l'analyseur syntaxique. Elles sont visibles lors de l'expérience pour présélectionner les équipes de patrons avec le plus de potentiel.

La présélection a exigé l'annotation de presque 500 phrases par trois juges. Seulement 22 phrases parmi celles-là ont été jugées pertinentes par les trois juges, unanimement. Or, parmi ces quelques phrases, l'extracteur automatique n'a pu extraire que onze éléments prédicatifs. Dès l'expérience de présélection, notre manière de capturer les verbes n'est efficace qu'à 50%. Nous avons examiné le comportement du système et nous avons noté trois types d'erreurs importantes.

Premièrement, les conjonctions ne sont pas supportées par nos patrons. C’est un exemple de la grande variété de phrases devant laquelle nos patrons sont trop peu nombreux, ou trop simples. Par exemple, la phrase pour *banana basket, things such as bananas, sugar cane, oranges, candles and joss sticks or other items to put into the bamboo baskets at the houses of the people they know* est rendue inutile par l’énumération d’éléments. Les mots *or* et *and* sont souvent impliqués. Ces conjonctions sont particulièrement dommageable quand ils séparent deux verbes valides, comme dans la phrase *Some other types of related bees produce and store honey*.

Deuxièmement, les erreurs de l’analyseur morphologique sont responsables de plusieurs verbes manqués. Le nom composé *data scientist* nous fournit un exemple avec la phrase *Computational linguistics allows scientists to parse huge amounts of data reliably and efficiently, creating the possibility for discoveries unlike any seen in most other approaches*. Le verbe (VB) *parse* est confondu avec un nom (NN) dans le groupe nominal (NP) *parse huge amounts*.

Troisièmement, retirer les guillemets des phrases fournirait quelques verbes supplémentaires. La phrase *As the first public bank to “offer accounts not directly convertible to coin”* contient le verbe *offer*, qui révèle la relation du nom composé *bank account*.

Nous avons aussi ignoré quelques prépositions, comme attendu. Les autres cas sont trop exceptionnels pour être traités individuellement, par des règles supplémentaires. Il est donc possible d’améliorer les performances d’un corpus avec une meilleure analyse syntaxique.

La quantité d’erreurs pour l’analyse de surface ne semble pas permettre de bons résultats avec une analyse syntaxique plus profonde, si cette dernière est basée sur les résultats parfois erronés de l’analyse de surface. L’extraction d’éléments prédicatifs à partir de petits corpus est sensible aux erreurs de l’analyseur. Les tentatives futures d’analyse syntaxique profonde devront donc prendre garde à l’accumulation des erreurs induites par les analyses précédentes du pipeline d’analyse.

6.4.2 Influence des vecteurs prédicatifs

Nous voulons trouver une métrique qui est corrélée avec le pourcentage de succès des expériences. Nous avons choisi trois expériences avec des taux de succès très différents : l’expérience de Nakov basée sur les vecteurs prédicatifs venant de Google, l’expérience de Yahoo! avec le plus haut taux de succès et ainsi que l’approche de Yahoo! qui a obtenu le pire score. Nous avons regroupé le nombre d’éléments prédicatifs dans le vecteur, la somme des fréquences d’apparition des éléments dans le vecteur et la somme des scores TF-IDF des éléments pour les trois expériences. Nous observons la somme, la moyenne et la médiane pour

tous les vecteurs prédictifs trouvés lors de l'identification des 212 noms composés donnés par Levi. Comme il est possible de le constater au tableau 6.10, il n'y a pas de corrélation apparente entre le taux de succès et le nombre d'éléments, leur fréquence ou leur score TF-IDF.

Tableau 6.10 Métriques pour les vecteurs d'éléments prédictifs trouvés pour les 212 noms composés

métrique	expérience	nb d'éléments distincts	fréquence	score TF-IDF
somme	Nakov (42%)	11827	84153	281834
	Yahoo! (33%)	3577	8640	31660
	Yahoo! (17%)	45068	158790	657083
moyenne	Nakov (42%)	56	397	1329
	Yahoo! (33%)	17	41	149
	Yahoo! (17%)	213	749	3099
médiane	Nakov (42%)	26	98	293
	Yahoo! (33%)	11	24	89
	Yahoo! (17%)	185	542	2274

Nous pouvons conclure que la nature des éléments composés a plus d'influence que leur nombre. Cette conclusion est renforcée par l'échec apparent des approches par synonyme.

6.4.3 Influence induite de la relation BE

L'analyse des erreurs de l'expérience témoin, avec Wikipédia pour corpus, révèle que la relation BE exerce une influence induite sur les réponses. Nous n'avons considéré dans l'analyse que les 179 noms composés pour lesquels il est possible de trouver un coefficient Dice plus grand que zéro à l'étape de comparaison. Le tableau 6.11 présente les succès divisés par relation. La deuxième colonne est le nombre de noms composés qui représentent une classe. La troisième colonne donne le pourcentage de mauvaise attribution qui est explicable si le nombre de représentants par phrase et le succès sont directement corrélés.

La relation IN semble être particulièrement difficile. Les noms composés de référence *weekend boredom*, *winter sports*, *surface tension* et *communist tenet* entre autres sont tous attribués à de mauvais prédicats.

Le tableau 6.12 fait la même analyse, mais tente de prédire les mauvaises attributions plutôt que les succès. La même expérience est utilisée avec les mêmes noms composés de référence, sauf qu'elle présente les mauvaises attributions, quand une relation donnée en réponse mène à une erreur.

Le ratio exact d'erreur pour la relation BE est de 49 sur 179 noms composés dont le plus

Tableau 6.11 Succès selon les relations sémantiques

Prédicat	Nombre	Représentation	Succès	Différence
BE	36	20%	53%	33%
ABOUT	14	8%	12%	4%
CAUSE ₁	7	4%	7%	3%
HAVE ₁	11	6%	7%	1%
MAKE ₁	4	2%	0%	-2%
USE	13	7%	5%	-3%
MAKE ₂	19	11%	7%	-4%
CAUSE ₂	12	7%	2%	-4%
HAVE ₂	13	7%	2%	-5%
FOR	15	8%	2%	-6%
FROM	17	9%	2%	-7%
IN	18	10%	0%	-10%
total	179	100 %	100%	0%

proche voisin a été trouvé (36%). Ce chiffre ne comprend pas les fois où la classe majoritaire BE a été attribuée parce qu’aucun nom composé n’était similaire.⁴ Si les erreurs étaient bien distribuées, 20% des erreurs devraient être causées par les 36 exemples de la relation BE, plutôt que 36%.

Notre système a clairement un penchant pour la relation BE. Plus de la moitié des succès sont pour des noms composés de relation BE, mais plus du tiers des erreurs aussi. Qu’est-ce qui entraîne le système à favoriser les représentants du prédicat BE ? De façon surprenante, aucune des mesures de la représentation des noms composés de référence en des nuages de points n’est corrélée au taux de succès ou au taux de mauvaise attribution. Ni la taille du nuage de points ni les différentes mesures de zone d’intersection ne peuvent prédire le degré de succès de la catégorie, ou sa présence dans les erreurs de relations.

Il est nécessaire de se pencher à nouveau sur la définition de la relation BE et mieux la définir, en modifiant les noms composés de référence qui causent le plus d’erreurs. Il serait même possible d’améliorer les résultats en éliminant complètement le prédicat BE.

6.5 Conclusions

Il n’est pas possible actuellement d’atteindre des taux de succès acceptables avec la méthode de Nakov pour les corpus autres que Google. Nous n’avons pas atteint le taux de succès de Nakov de 42%. Nous tirons de nos expériences trois conclusions.

4. Ces cas adviennent quand aucun des éléments prédicatifs trouvés dans le corpus n’est présent dans les vecteurs de référence. Le coefficient Dice est nul.

Tableau 6.12 Mauvaises attributions selon les relations sémantiques

Prédicat	Nombre	Représentation	Mauvaises attributions	Différence
BE	36	20%	36%	16%
HAVE ₁	11	6%	13%	6%
MAKE ₁	4	2%	6%	4%
CAUSE ₁	7	4%	7%	3%
MAKE ₂	19	11%	10%	0%
USE	13	7%	5%	-2%
FOR	15	8%	6%	-2%
ABOUT	14	8%	5%	-3%
HAVE ₂	13	7%	4%	-4%
CAUSE ₂	12	7%	1%	-5%
IN	18	10%	4%	-6%
FROM	17	9%	3%	-7%
total	179	100 %	100%	0%

Premièrement, le succès de Yahoo!, qui a le meilleur taux de catégorisation avec 33% laisse penser qu'il n'est pas encore envisageable de remplacer le potentiel d'un engin de recherche web comme corpus. Malgré les ressources exigées, une ressource web semble être la seule façon de trouver des documents pour les noms composés rares. Si trouver des éléments prédicatifs pour *love song* est facile, c'est un tout autre défi de fournir du contexte pour *chop suey* ou *weekend boredom*.

Toutefois, plusieurs de nos résultats semblent encourager l'utilisation des corpus alternatifs. D'abord, les synonymes triés par un juge utilisé en combinaison avec Yahoo! ont augmenté massivement la quantité d'éléments dans le vecteur du corpus, selon toutes mesures, en atteignant un désastreux 17% de noms composés bien catégorisés. Que la qualité des éléments du vecteur soit plus importante que la quantité est une conclusion encourageante pour les corpus alternatifs. Ce point de vue s'appuie aussi sur les mauvais résultats des patrons favorisant le rappel comparativement aux patrons de Nakov. Le surplus d'éléments prédicatifs n'a pas amené de meilleurs résultats. Avec les améliorations pour l'analyse des phrases du corpus soulignées à la section 6.4.1, nous pouvons envisager de meilleurs résultats avec le corpus de Wikipédia. Combiner les corpus reste une avenue avec beaucoup de potentiel.

Deuxièmement, les expériences nous démontrent l'importance des vecteurs de références dans la méthode de Nakov. Nous avons observé à la section 6.3.2 que les synonymes peuvent nuire si les vecteurs de référence sont trop petits pour être correctement comparés au vecteur prédicatif du corpus. Le petit nombre d'éléments de référence nuit à la stabilité des réponses. Nous avons observé à la section 6.3.4 qu'un seul verbe supplémentaire dans le vecteur pré-

dicatif du corpus entraîne un grand changement dans l’ordonnancement par similarité. Il ne faut pas oublier que les noms composés de référence favorisent le prédicat BE. 36% des mauvaises attributions sont causées par ce prédicat. Les vecteurs de référence pour ce prédicat sont suspects et devraient être les premiers à être rectifiés dans les tentatives futures. Toute implémentation future de l’approche de Nakov devrait donc porter une attention particulière à la qualité des vecteurs de référence et tenter d’augmenter leur nombre.

Troisièmement et dernièrement, changer la formule de comparaison Dice pourrait entraîner une amélioration de résultat. Une équation qui inclut une pondération partielle pour des verbes identiques, mais des particules différentes augmenterait les scores qui restent nuls pour la majorité des noms composés de référence. Il est important de mettre à contribution tous les éléments prédicatifs extraits du corpus, au moins en partie. Dans le cas présent, avec le coefficient Dice, seuls ceux qui sont identiques à un élément prédicatif de référence contribuent à la comparaison. Par ailleurs, si la comparaison incluait une normalisation des scores TF-IDF par vecteur de référence et une multiplication par les scores du vecteur du corpus, la difficulté de comparer petits vecteurs de référence et grand vecteur du corpus serait réglée.

Malgré les possibilités d’amélioration mentionnées plus haut, nous ne pouvons nier que l’approche implémentée n’arrive qu’à la cheville des autres approches de la littérature. Il nous faut conclure qu’actuellement, l’implémentation simple de l’approche de Nakov n’est pas suffisante pour la catégorisation de noms composés parmi 12 relations implicites. Les trois approches notables de la tâche 4 de SemEval-2013 (Hendrickx et al., 2013) incluent IIITH (Surtani et al., 2013), SFS (Versley, 2013) et MELODI (Cruys and Muller, 2013).

Le système de l’université IIITH (International Institute of Information Technology, Hyderabad) (Surtani et al., 2013), extrait des paraphrases des n-grammes de Google et des corpus ANC et BNC. Il utilise les verbes du corpus qui prennent la place des verbes connus dans un ensemble de paraphrases initiales. Si les paraphrases ne sont pas suffisamment nombreuses, l’algorithme se base uniquement sur les paraphrases avec préposition seule, comme *party in the garden*.

Versley utilise une approche par paraphrase similaire à la nôtre, avec les n-grammes de Google et les paraphrases des juges humains fournies par Nakov. Différents motifs de paraphrases, incluant les prépositions et les verbes courants, sont récoltés et jugés par une combinaison de dix mesures de similarité.

L’approche de Cruys utilise une matrice de co-occurrence des noms des noms composés par les mots trouvés dans le contexte. La rangée du premier nom est multipliée par la rangée du second nom. Le score résultat est utilisé dans un modèle d’apprentissage supervisé.

Nous espérons que l'implémentation de ces autres approches réussisse mieux à identifier avec succès la relation implicite dans les noms composés.

CHAPITRE 7 CONCLUSION ET TRAVAUX FUTURS

Au fil de ce mémoire, nous avons voulu révéler la relation sémantique implicite au cœur des noms composés. Nous avons expliqué la problématique des noms composés dans le contexte du traitement automatique des langues naturelles. Nous avons détaillé tout un éventail d'approches, utilisant les hyperonymes de Wordnet, un réseau neuronal ou une machine à vecteur de support. Nous avons choisi de reproduire l'approche de Nakov autant qu'il était possible sans l'accès au moteur de recherche de Google. Pour le remplacer, nous avons choisi les articles anglophones de Wikipédia, les n-grammes de Google tirés du web et les moteurs de recherche de Yahoo et de Faroo.

L'approche se décompose en quatre étapes. La première est l'extraction de phrases pertinentes à partir du corpus. La deuxième est l'analyse syntaxique à l'aide de patrons lexicaux pour trouver des éléments prédicatifs, ces verbes et prépositions qui révèlent la relation entre les deux noms du nom composé. La troisième est la comparaison avec des noms composés dont la relation est connue. La quatrième étape est la déduction de la relation sémantique finale. Nous avons refait l'expérience avec différentes versions de chaque étape, en ajoutant différentes sources de synonymes pour interroger nos différents corpus, en variant la composition du groupe de patrons lexicaux utilisés dans l'analyse, en réorganisant la façon de représenter les noms composés connus pour la comparaison.

Là où Nakov atteint un taux de succès de 43%, nous n'obtenons qu'un taux de succès de 33%, avec le corpus de Yahoo. Nous n'avons pas pu dépasser ce score à l'aide des autres corpus, malgré nos nombreuses modifications au protocole. Notre meilleure tentative pour les corpus hors-ligne atteint le taux de succès de 28%, avec Wikipédia. Malheureusement, la version du protocole qui emploie des synonymes choisis manuellement a récolté plus de phrases pour décrire le contexte du nom composé, mais ce gain d'information n'a pas mené à de meilleurs résultats. L'inversion et la révision des éléments prédicatifs pour mieux les comparer aux éléments prédicatifs tirés du corpus ont autant nui aux résultats qu'elles ne les ont améliorés.

Le succès apparent de Yahoo semble être dû à l'extraction de paraphrases pour les noms composés inusités, où les autres corpus ne trouvent aucune paraphrase. Cependant, après un certain nombre de paraphrase extraites, c'est leur qualité qui a un impact, d'après les résultats de l'ajout de noms composés synonymes et de l'extraction avec des patrons plus permissifs. Il est donc toujours possible d'envisager de bons résultats avec les corpus de qualité disponibles hors-ligne, plus faciles à accéder. Une amélioration de l'analyse syntaxique des phrases et la

combinaison possible des corpus encourage la poursuite de l'utilisation de ces corpus.

Par ailleurs, nos résultats soulignent l'importance des vecteurs prédictifs utilisés comme référence. Ils sont déterminants pour l'amélioration de l'approche. Enfin, changer la façon de comparer les listes d'éléments du corpus avec les listes de référence pour tenir compte des paraphrases partiellement similaires exploitera beaucoup mieux les données accumulées.

Si ces pistes n'améliorent pas les faibles résultats de l'approche, il est possible de s'inspirer des paradigmes utilisés par les gagnants de la tâche de paraphrase de noms composés, à l'atelier de SemEval-2013 (Hendrickx et al., 2013).

Passer de notre prototype de recherche à l'application avec les descriptions d'inventaire pour les produits traversant la frontière sera difficile en raison de plusieurs limitations, au-delà des faibles résultats. Le temps de calcul le plus court, après mise en cache des résultats du corpus, prend une heure avec Faroo. Le grand corpus de Yahoo peut prendre jusqu'à 5 jours de calcul, si tous les synonymes sont inclus. La majorité du temps est employée à annoter chaque mot des phrases avec son syntagme et son rôle dans la phrase et à comparer la phrase annotée à tous les patrons. L'analyse utilise aussi 4 Go de mémoire vive et jusqu'à 8 Go si de très longues phrases sont analysées. Il manque à notre méthode un module qui reconnaît les noms composés formés avec un verbe, comme *software developer*. Assigner une catégorie sémantique au nom composé qui se paraphrase par *one that develops software* donne des résultats approximatifs, alors que reconnaître ces cas ne devrait pas poser de grandes difficultés.

Pour le moment, l'analyse des descriptions d'inventaire de l'entreprise 3CE ne pourra pas exploiter la relation implicite des noms composés. Toutefois, dans les recherches futures, au-delà de quelques correctifs, nous envisageons trois avenues d'innovation différentes.

La première avenue, c'est d'essayer différentes manières de catégoriser les relations. Utiliser les mêmes catégories que celles employées lors de la conférence SemEval diminuera le nombre de catégories possibles et permettra de nous comparer à beaucoup d'autres approches. Par ailleurs, nous avons souligné l'importance dans l'approche d'avoir beaucoup de noms composés connus et pour chacun, une grande liste d'éléments prédictifs pour faciliter la comparaison. Pour augmenter ce facteur, il faut remplacer les exemples et relations de Levi par les relations de Barker & Szpakowicz (Barker and Szpakowicz, 1998), comme l'a fait Nakov (Nakov and Hearst, 2008), et utiliser la liste de 1000 noms composés d'entraînement de Kim, chacun annoté avec l'une des 20 relations de Barker. Bien d'autres ensembles de relations et corpus sont disponibles.

La deuxième principale ligne d'amélioration est la plus facile à faire et pourrait augmenter

les résultats. Combiner les vecteurs prédictifs provenant de différents corpus permettrait enfin d'avoir autant de résultats qu'avec le corpus initial de Google. Avec différentes sources, il serait possible de ne conserver que celle qui fournissent les phrases de contexte les plus révélatrices de la relation pour conserver la qualité des éléments prédictifs en même d'en d'augmenter leur nombre. Il faudra s'assurer que les dernières étapes de la démarche utilisent efficacement les longs vecteurs prédictifs, pour éviter de rencontrer les mêmes obstacles que lors de l'utilisation des synonymes.

La troisième manière d'innover serait de faire une analyse syntaxique profonde de la phrase pour trouver sa structure. Une analyse syntaxique présente la phrase comme un arbre, dans lequel les mots sont des feuilles, qui sont organisées en groupes de plus en plus gros, jusqu'à former la phrase complète à la souche. Cette avenue d'exploration continue dans le sens des approches de patrons sévères, qui ne choisissent que les verbes dans des structures de phrases très précises, pour augmenter la pertinence du verbe. L'analyse syntaxique profonde connaîtra mieux la phrase et pourra peut-être mieux choisir les verbes, avec des patrons qui tiennent compte non seulement des syntagmes, mais aussi des liens avec les autres groupes de mots. Il faudra faire attention aux erreurs que l'analyse en arbre introduit dans le processus. Si l'analyse syntaxique simple a causé plusieurs erreurs, l'analyse syntaxique profonde, qui est basée sur les résultats de l'analyse syntaxique simple, en causera plus. Plus les analyseurs de phrases, dont l'analyseur de Stanford (de Marneffe et al.), s'améliorent, plus le potentiel de cette avenue d'exploration augmentera.

RÉFÉRENCES

- K. Barker et S. Szpakowicz, “Semi-Automatic Recognition of Noun Modifier Relationships,” dans *Proceedings of the 36th Annual Meeting of the Association for Computational Linguistics and 17th International Conference on Computational Linguistics - Volume 1*, Montréal, 1998, pp. 96–102. [En ligne]. Disponible : <http://dl.acm.org/citation.cfm?id=980451.980862><http://aclweb.org/anthology/P/P98/P98-1015.pdf><http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.27.6484&rep=rep1&type=pdf>
- B. Beamer, S. Bhat, B. Chee, A. Fister, A. Rozovskaya, et R. Girju, “UIUC : A Knowledge-Rich Approach to Identifying Semantic Relations between Nominals,” dans *Proceedings of the 4th International Workshop on Semantic Evaluations (SemEval-1)*, Prague, 2007, pp. 386–389.
- C. Butnariu et T. Veale, “A Concept-Centered Approach to Noun-Compound Interpretation,” dans *International Conference on Computational Linguistics - COLING*, 2008, pp. 81–88. [En ligne]. Disponible : http://afflatus.ucd.ie/Papers/nounCompounds_Coling08.pdf<http://afflatus.ucd.ie/Papers/>
- S. Coulson, *Semantic leaps : Frame-Shifting and Conceptual Blending in Meaning Construction*. Cambridge : Cambridge University Press, 2001.
- T. V. D. Cruys et P. Muller, “MELODI : A Supervised Distributional Approach for Free Paraphrasing of Noun Compounds,” dans *Second Joint Conference on Lexical and Computational Semantics (*SEM), Volume 2 : Proceedings of the Seventh International Workshop on Semantic Evaluation (SemEval 2013)*, vol. 2, no. SemEval. Atlanta : Association for Computational Linguistics, 2013, pp. 144–147. [En ligne]. Disponible : <http://www.aclweb.org/anthology/S13-2026>
- M.-C. de Marneffe, B. Maccartney, et C. D. Manning, “Generating Typed Dependency Parses from Phrase Structure Parses,” dans *LREC 2006*, pp. 449–454. [En ligne]. Disponible : <http://nlp.stanford.edu/software/lex-parser.shtml>
- L. R. Dice, “Measures of the Amount of Ecologic Association Between Species,” *Ecology*, vol. 26, no. 3, pp. 297–302, Jul. 1945. [En ligne]. Disponible : <http://dx.doi.org/10.2307/1932409>
- P. A. Downing, “On the Creation and Use of English Compound Nouns,” *Language*, vol. 53, no. 4, pp. 810–842, 1977. [En ligne]. Disponible : <http://www.jstor.org/discover/10.2307/412913?uid=3739464&uid=2&uid=3737720&uid=4&sid=21104582419083><http://www.doc88.com/p-1901926515092.html>

- R. Girju, B. Beamer, A. Rozovskaya, A. Fister, et S. Bhat, “A Knowledge-Rich Approach to Identifying Semantic Relations between Nominals,” *Information Processing & Management*, vol. 46, no. 5, pp. 589–610, Sep. 2009. [En ligne]. Disponible : <http://linkinghub.elsevier.com/retrieve/pii/S0306457309001046>
- R. Girju, “Experiments with an Annotation Scheme for a Knowledge-rich Noun Phrase Interpretation System,” dans *Proceedings of the Linguistic Annotation Workshop*, Association for Computational Linguistics, éd., no. June, Prague, 2007, pp. 168–175.
- R. Girju, P. Nakov, V. Nastase, S. Szpakowicz, P. Turney, et D. Yuret, “SemEval-2007 Task 04 : Classification of Semantic Relations between Nominals,” no. June, pp. 13–18, 2007. [En ligne]. Disponible : <http://hlt.fbk.eu/people/profile/nastase>
- I. Hendrickx, Z. Kozareva, P. Nakov, D. Ó Séaghdha, S. Szpakowicz, et T. Veale, “SemEval-2013 Task 4 : Free Paraphrases of Noun Compounds,” dans *Proceedings of the Second Joint Conference on Lexical and Computational Semantics (*SEM'13)*, vol. 2, no. SemEval, Atlanta, 2013, pp. 138–143. [En ligne]. Disponible : http://people.ischool.berkeley.edu/~nakov/selected_papers_list/
- S. N. Kim et T. Baldwin, “Automatic Interpretation of Noun Compounds Using WordNet Similarity,” dans *Natural Language Processing – IJCNLP 2005*, sér. Lecture Notes in Computer Science, R. Dale, K.-F. Wong, J. Su, et O. Kwong, éd., vol. 3651. Springer Berlin Heidelberg, 2005, pp. 945–956. [En ligne]. Disponible : http://dx.doi.org/10.1007/11562214_82
- , “Interpreting Semantic Relations in Noun Compounds via Verb Semantics,” dans *Proceedings of the 44th Annual Meeting of the Association for Computational Linguistics and 21st International Conference on Computational Linguistics*, Sydney, 2006, pp. 491–498.
- S. N. Kim et P. Nakov, “Large-Scale Noun Compound Interpretation Using Bootstrapping and the Web as a Corpus,” *EMNLP '11 Proceedings of the Conference on Empirical Methods in Natural Language Processing*, pp. 648–658, 2011. [En ligne]. Disponible : <http://dl.acm.org/citation.cfm?id=2145508>
- M. Lauer, “Designing Statistical Language Learners : Experiments on Noun Compounds,” Thèse de doctorat, Macquarie University, 1995.
- R. Leonard, *The Interpretation of English Noun Sequences on the Computer*. Amsterdam : Elsevier Science & Technology, 1984.
- J. N. Levi, *The Syntax and Semantics of Complex Nominals*. New York : Academic Press, Inc, 1978.
- D. Moldovan, A. Badulescu, M. Tatu, D. Antohe, et R. Girju, “Models for the Semantic Classification of Noun Phrases,” dans *HLT-NAACL 2004 : Workshop*

on *Computational Lexical Semantics*, no. 2, 2004, pp. 60–67. [En ligne]. Disponible : <http://www.cs.brandeis.edu/~marc/misc/proceedings/hlt-naacl-2004/CLS/index.html><http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.2.2528>

P. Nakov, “Paraphrasing Verbs for Noun Compound Interpretation,” dans *Proceedings of the LREC 2008 Workshop : Towards a Shared Task for Multiword Expressions (MWE 2008)*, Marrakech, 2008. [En ligne]. Disponible : http://link.springer.com/chapter/10.1007/978-3-540-85776-1_10<http://biotext.berkeley.edu/papers/mwe2008.pdf>

——, “On the interpretation of noun compounds : Syntax, semantics, and entailment,” *Natural Language Engineering*, vol. 19, no. 03, pp. 291–330, Mai 2013. [En ligne]. Disponible : http://www.journals.cambridge.org/abstract_S1351324913000065<http://www.qcri.qa/app/media/2011/>

P. Nakov et M. Hearst, “Semantic Interpretation of Noun Compounds Using Verbal and Other Paraphrases,” *ACM Transactions on Speech and Language Processing*, vol. 0, no. 0, 2013. [En ligne]. Disponible : http://people.ischool.berkeley.edu/~nakov/selected_papers_list/TSLP2013.pdf

P. Nakov et M. A. Hearst, “Solving Relational Similarity Problems Using the Web as a Corpus,” dans *Proceedings of the ACL’08 : HLT*. Columbus, OH : Association for Computational Linguistic, 2008, pp. 452–460. [En ligne]. Disponible : <http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.143.4811>

P. I. Nakov et M. A. Hearst, “UCB : System Description for SemEval Task # 4,” dans *Proceedings of the Fourth International Workshop on Semantic Evaluations (SemEval-2007)*, vol. 94720, no. June. Prague : Association for Computational Linguistics, 2007, pp. 366–369. [En ligne]. Disponible : <http://www.aclweb.org/anthology/S/S07/S07-1080><http://aclweb.org/anthology/S/S07/>

V. Nastase et S. Szpakowicz, “Exploring noun-modifier semantic relations,” dans *Fifth International Workshop on Computational Semantics*, Potsdam, 2003, pp. 285–301.

V. Nastase, J. Sayyad-shirabad, M. Sokolova, et S. Szpakowicz, “Learning Noun-Modifier Semantic Relations with Corpus-based and WordNet-based Features,” dans *Proceedings of the 21st National Conference on Artificial Intelligence*. Boston : Intelligence, American Association for Artificial, 2006, pp. 781–787. [En ligne]. Disponible : <http://corevocab.com/?page=SkE/Biblio><http://www.h-its.org/english/homes/nastase/Publications/AAAI0613NastaseV.pdf><http://www.aaai.org/Library/AAAI/aaai06contents.php#technical>

V. Nastase, P. Nakov, D. Ó Séaghdha, et S. Szpakowicz, *Semantic Relations Between Nominals*, G. Hirst, éd. Toronto : Morgan & Claypool, 2013.

- D. Ó Séaghdha, “Learning compound noun semantics,” Thèse de doctorat, University of Cambridge, 2008.
- S. Patwardhan, S. Banerjee, et T. Pedersen, “Using Measures of Semantic Relatedness for Word Sense Disambiguation,” dans *Proceedings of the Fourth International Conference on Intelligent Text Processing and Computational Linguistics*, vol. 4, Mexico, 2003, pp. 241–257. [En ligne]. Disponible : <http://www.d.umn.edu/~tpederse/similarity-pubs.html>
- Princeton University, “About WordNet.” <http://wordnet.princeton.edu>, 2010. [En ligne]. Disponible : <http://wordnet.princeton.edu>
- B. Rink et S. Harabagiu, “UTD : Classifying Semantic Relations by Combining Lexical and Semantic Resources,” dans *Proceedings of the 5th International Workshop on Semantic Evaluation*, Uppsala, 2010, pp. 256–259. [En ligne]. Disponible : <http://www.aclweb.org/anthology/S10-1057>
- B. Rosario et M. Hearst, “Classifying the Semantic Relations in Noun Compounds via a Domain-Specific Lexical Hierarchy,” dans *Proceedings of 2001 Conference on Empirical Methods in Natural Language Processing*, L. Lee et D. Harman, édés., Pittsburgh, 2001, pp. 82–90. [En ligne]. Disponible : <http://berkeley.intel-research.net/~broshario/papers.html><http://www.cs.cornell.edu/home/llee/emnlp/proceedings.html><http://www.cs.cornell.edu/home/llee/emnlp/papers/rosario.pdf>
- B. Santorini, “Part-of-Speech Tagging Guidelines for the Penn Treebank Project (3rd Revision),” University of Pennsylvania Department of Computer and Information Science, Philadelphie, Rapp. tech., 1990. [En ligne]. Disponible : http://repository.upenn.edu/cgi/viewcontent.cgi?article=1603&context=cis_reportshttp://repository.upenn.edu/cis_reports/570/
- T. Sørensen, “A method of establishing groups of equal amplitude in plant sociology based on similarity of species and its application to analyses of the vegetation on danish commons,” *Biologiske Skrifter / Kongelige Danske Videnskabernes Selskab*, vol. 5, no. 4, pp. 1–34, 1948.
- N. Surtani, A. Batra, U. Ghosh, et S. Paul, “IIITH : A Corpus-Driven Co-occurrence Based Probabilistic Model for Noun Compound Paraphrasing,” dans *Second Joint Conference on Lexical and Computational Semantics (*SEM), Volume 2 : Proceedings of the Seventh International Workshop on Semantic Evaluation (SemEval 2013)*, vol. 2. Atlanta : Association for Computational Linguistics, 2013, pp. 153–157. [En ligne]. Disponible : <http://www.aclweb.org/anthology/S13-2028>
- P. D. Turney, “Measuring Semantic Similarity by Latent Relational Analysis,” dans *Proceedings of the Nineteenth International Joint Conference on Artificial Intelligence (IJCAI-05)*, Edinburgh, Scotland, 2005, pp. 1136–1141.

L. Vanderwende, “Algorithm for the automatic interpretation of noun sequences,” dans *Proceedings of the 15th International Conference on Computational Linguistics*, Kyoto, 1994, pp. 782–788.

Y. Versley, “SFS-TUE : Compound Paraphrasing with a Language Model and Discriminative Reranking,” dans *Second Joint Conference on Lexical and Computational Semantics (*SEM), Volume 2 : Proceedings of the Seventh International Workshop on Semantic Evaluation (SemEval 2013)*, Atlanta, 2013, pp. 148–152. [En ligne]. Disponible : http://www.cs.york.ac.uk/semeval-2013/accepted/42_Paper.pdf

K. Zimmer, “Some General Observations about Nominal Compounds,” *Working Papers on Language Universals*, vol. 5, pp. 1–21, 1971.

ANNEXE A Noms composés de Levi

- | | |
|--|---|
| 1. abortion problem : <i>about</i> | 28. child actor : <i>be</i> |
| 2. abortion vote : <i>about</i> | 29. childhood dreams : <i>in</i> |
| 3. accident weather : <i>cause₁</i> | 30. chocolate bar : <i>make₂</i> |
| 4. adolescent turmoil : <i>in</i> | 31. cigarette burn : <i>cause₂</i> |
| 5. adventure story : <i>about</i> | 32. cigarette war : <i>about</i> |
| 6. air pressure : <i>cause₂</i> | 33. citizen soldier : <i>be</i> |
| 7. alligator leather : <i>from</i> | 34. city folk : <i>in</i> |
| 8. apple cake : <i>have₁</i> | 35. city wall : <i>have₂</i> |
| 9. apple core : <i>have₂</i> | 36. coal dust : <i>from</i> |
| 10. arms budget : <i>for</i> | 37. coffee nerves : <i>cause₂</i> |
| 11. automobile plant : <i>for</i> | 38. coke machine : <i>for</i> |
| 12. bacon grease : <i>from</i> | 39. college town : <i>have₁</i> |
| 13. basketball season : <i>for</i> | 40. collie dog : <i>be</i> |
| 14. bear country : <i>have₁</i> | 41. color television : <i>have₁</i> |
| 15. beehive hairdo : <i>be</i> | 42. communist tenet : <i>in</i> |
| 16. birth pains : <i>cause₂</i> | 43. company assets : <i>have₂</i> |
| 17. blanket excuse : <i>be</i> | 44. concrete desert : <i>make₂</i> |
| 18. bronze statue : <i>make₂</i> | 45. concussion force : <i>cause₁</i> |
| 19. budget speech : <i>about</i> | 46. cooking utensils : <i>for</i> |
| 20. bull ring : <i>for</i> | 47. copper coins : <i>make₂</i> |
| 21. cable network : <i>make₂</i> | 48. coriander curry : <i>have₁</i> |
| 22. cactus plant : <i>be</i> | 49. country butter : <i>from</i> |
| 23. candy cigarette : <i>make₂</i> | 50. country visitors : <i>from</i> |
| 24. cane sugar : <i>from</i> | 51. cream sauce : <i>have₁</i> |
| 25. canine companion : <i>be</i> | 52. daisy chains : <i>make₂</i> |
| 26. cash basis : <i>be</i> | 53. deficiency disease : <i>cause₂</i> |
| 27. ceiling price : <i>be</i> | 54. desert rat : <i>in</i> |
| | 55. disaster flick : <i>about</i> |

56. disease germ : *cause*₁
57. drug deaths : *cause*₂
58. enemy strength : *have*₂
59. energy emergency : *about*
60. evening hours : *in*
61. extension ladder : *have*₁
62. faith cure : *use*
63. family antiques : *have*₂
64. family problems : *in*
65. farm boy : *from*
66. fatigue headache : *cause*₂
67. field mouse : *in*
68. financing dilemma : *about*
69. finger cymbals : *use*
70. finger lakes : *be*
71. fish scales : *have*₂
72. flounder fish : *be*
73. flu virus : *cause*₁
74. frog man : *be*
75. fruit tree : *have*₁
76. future shock : *cause*₂
77. garter snake : *be*
78. gas stove : *use*
79. glass eye : *make*₂
80. government employment : *in*
81. government land : *have*₂
82. grain alcohol : *from*
83. growth hormone : *cause*₁
84. gutter language : *from*
85. hairpin turn : *be*
86. hand brake : *use*
87. handlebar mustache : *be*
88. head noun : *be*
89. headache pills : *for*
90. heart design : *be*
91. heat rash : *cause*₂
92. hermit crab : *be*
93. history conference : *about*
94. home remedy : *from*
95. honey bee : *make*₁
96. horse doctor : *for*
97. hydrogen bomb : *use*
98. immigrant minority : *make*₂
99. infant colonies : *be*
100. job tension : *cause*₂
101. juvenile court : *for*
102. kennel puppies : *from*
103. lace handkerchief : *have*₁
104. language riots : *about*
105. laugh wrinkles : *cause*₂
106. lemon peel : *have*₂
107. lightning rod : *for*
108. lion cub : *be*
109. love song : *about*
110. machine translation : *use*
111. marine life : *in*
112. midnight snack : *in*
113. milieu therapy : *use*
114. mining engineer : *for*
115. morning prayers : *in*
116. morphology lecture : *about*
117. moth hole : *cause*₂

118. mother church : *be*
119. mountain lodge : *in*
120. mountain range : *make₂*
121. movement schisms : *in*
122. murder charge : *be*
123. music box : *make₁*
124. neighborhood bars : *in*
125. night flight : *in*
126. nose drops : *for*
127. novelty item : *be*
128. nut bread : *have₁*
129. office friendships : *in*
130. oil crisis : *about*
131. oil well : *for*
132. olive oil : *from*
133. onion tears : *cause₂*
134. paper money : *make₂*
135. parent organization : *be*
136. party members : *have₂*
137. peanut butter : *from*
138. pedal extremities : *be*
139. people power : *have₂*
140. pet families : *have₁*
141. pet spray : *for*
142. pet theory : *be*
143. phantom limb : *be*
144. picture album : *for*
145. picture book : *have₁*
146. pine tree : *be*
147. plant food : *for*
148. plastic toys : *make₂*
149. plum wine : *from*
150. pole height : *have₂*
151. policy matters : *about*
152. pork suet : *from*
153. pot high : *cause₂*
154. pressure cooker : *use*
155. price dispute : *about*
156. queen bee : *be*
157. radio communication : *use*
158. rice paper : *from*
159. rye whiskey : *from*
160. salt lake : *have₁*
161. sand dune : *make₂*
162. sap tree : *make₁*
163. satellite nation : *be*
164. sea breeze : *from*
165. servant girl : *be*
166. sex scandal : *about*
167. shock treatment : *use*
168. sister node : *be*
169. smoke signals : *use*
170. snow blindness : *cause₂*
171. sob story : *cause₁*
172. soldier ant : *be*
173. song bird : *make₁*
174. sports activities : *be*
175. sports magazine : *about*
176. star shape : *be*
177. starvation diet : *use*
178. steam iron : *use*
179. steel helmet : *make₂*

- | | |
|---|---|
| 180. stone tools : <i>make</i> ₂ | 197. tire rim : <i>have</i> ₂ |
| 181. store clothes : <i>from</i> | 198. tobacco ash : <i>from</i> |
| 182. student committee : <i>make</i> ₂ | 199. vacuum cleaner : <i>use</i> |
| 183. student friends : <i>be</i> | 200. vapor lock : <i>cause</i> ₂ |
| 184. student power : <i>have</i> ₂ | 201. vegetable soup : <i>have</i> ₁ |
| 185. student problems : <i>have</i> ₂ | 202. voice vote : <i>use</i> |
| 186. sugar cube : <i>make</i> ₂ | 203. warrior caste : <i>make</i> ₂ (non utilisé) |
| 187. summer months : <i>in</i> | 204. wastebasket category : <i>be</i> |
| 188. summer travels : <i>in</i> | 205. water drop : <i>make</i> ₂ |
| 189. surface tension : <i>in</i> | 206. water mark : <i>cause</i> ₂ |
| 190. suspense film : <i>cause</i> ₁ | 207. water wheel : <i>use</i> |
| 191. tape measure : <i>be</i> | 208. weekend boredom : <i>in</i> |
| 192. target structure : <i>be</i> | 209. winter season : <i>be</i> |
| 193. tax law : <i>about</i> | 210. winter sports : <i>in</i> |
| 194. teaching profession : <i>be</i> | 211. women professors : <i>be</i> |
| 195. tear gas : <i>cause</i> ₁ | 212. wood shavings : <i>from</i> |
| 196. testtube baby : <i>from</i> | 213. worker teams : <i>make</i> ₂ |

ANNEXE B Noms composés de Levi recatégorisés

- | | |
|--|---|
| 1. abortion problem : <i>about</i> | 28. child actor : <i>be</i> |
| 2. abortion vote : <i>about</i> | 29. childhood dreams : <i>from, in</i> |
| 3. accident weather : <i>cause₁</i> | 30. chocolate bar : <i>be, make₂</i> |
| 4. adolescent turmoil : <i>have₂, in</i> | 31. cigarette burn : <i>cause₂</i> |
| 5. adventure story : <i>about, have₁</i> | 32. cigarette war : <i>about</i> |
| 6. air pressure : <i>cause₂, have₂</i> | 33. citizen soldier : <i>be</i> |
| 7. alligator leather : <i>from</i> | 34. city folk : <i>from, in</i> |
| 8. apple cake : <i>have₁</i> | 35. city wall : <i>have₂</i> |
| 9. apple core : <i>have₂</i> | 36. coal dust : <i>from, make₂</i> |
| 10. arms budget : <i>for</i> | 37. coffee nerves : <i>cause₂</i> |
| 11. automobile plant : <i>make₁</i> | 38. coke machine : <i>have₁</i> |
| 12. bacon grease : <i>from, have₂</i> | 39. college town : <i>have₁</i> |
| 13. basketball season : <i>for, have₂</i> | 40. collie dog : <i>be</i> |
| 14. bear country : <i>have₁</i> | 41. color television : <i>have₁, use</i> |
| 15. beehive hairdo : <i>be</i> | 42. communist tenet : <i>have₂</i> |
| 16. birth pains : <i>cause₂, in</i> | 43. company assets : <i>have₂</i> |
| 17. blanket excuse : <i>be</i> | 44. concrete desert : <i>make₂</i> |
| 18. bronze statue : <i>make₂</i> | 45. concussion force : <i>cause₁</i> |
| 19. budget speech : <i>about</i> | 46. cooking utensils : <i>for</i> |
| 20. bull ring : <i>for</i> | 47. copper coins : <i>make₂</i> |
| 21. cable network : <i>make₂, use</i> | 48. coriander curry : <i>have₁</i> |
| 22. cactus plant : <i>be</i> | 49. country butter : <i>from</i> |
| 23. candy cigarette : <i>be</i> | 50. country visitors : <i>from, in</i> |
| 24. cane sugar : <i>from</i> | 51. cream sauce : <i>have₁, make₂</i> |
| 25. canine companion : <i>be</i> | 52. daisy chains : <i>make₂</i> |
| 26. cash basis : <i>be, use</i> | 53. deficiency disease : <i>cause₂</i> |
| 27. ceiling price : <i>be</i> | 54. desert rat : <i>in</i> |
| | 55. disaster flick : <i>about, have₁</i> |

56. disease germ : *cause*₁
57. drug deaths : *cause*₂
58. enemy strength : *have*₂
59. energy emergency : *about*
60. evening hours : *in*
61. extension ladder : *be*
62. faith cure : *use*
63. family antiques : *from, have*₂
64. family problems : *about, have*₂, *in*
65. farm boy : *from*
66. fatigue headache : *cause*₂
67. field mouse : *in*
68. financing dilemma : *about*
69. finger cymbals : *use*
70. finger lakes : *be*
71. fish scales : *have*₂
72. flounder fish : *be*
73. flu virus : *cause*₁
74. frog man : *be*
75. fruit tree : *have*₁, *make*₁
76. future shock : *cause*₂
77. garter snake : *be*
78. gas stove : *use*
79. glass eye : *make*₂
80. government employment : *for, in*
81. government land : *have*₂
82. grain alcohol : *from*
83. growth hormone : *cause*₁, *for*
84. gutter language : *from*
85. hairpin turn : *be*
86. hand brake : *use*
87. handlebar mustache : *be*
88. head noun : *be*
89. headache pills : *for*
90. heart design : *be*
91. heat rash : *cause*₂
92. hermit crab : *be*
93. history conference : *about*
94. home remedy : *from*
95. honey bee : *make*₁
96. horse doctor : *for*
97. hydrogen bomb : *use*
98. immigrant minority : *make*₂
99. infant colonies : *be*
100. job tension : *cause*₂
101. juvenile court : *for*
102. kennel puppies : *from*
103. lace handkerchief : *have*₁, *make*₂
104. language riots : *about*
105. laugh wrinkles : *cause*₂
106. lemon peel : *from, have*₂
107. lightning rod : *for*
108. lion cub : *be*
109. love song : *about*
110. machine translation : *make*₂
111. marine life : *in*
112. midnight snack : *in*
113. milieu therapy : *use*
114. mining engineer : *about, for*
115. morning prayers : *in*
116. morphology lecture : *about*
117. moth hole : *cause*₂

118. mother church : *be*
119. mountain lodge : *in*
120. mountain range : *make₂*
121. movement schisms : *in*
122. murder charge : *about, for*
123. music box : *make₁*
124. neighborhood bars : *in*
125. night flight : *in*
126. nose drops : *for*
127. novelty item : *be*
128. nut bread : *have₁*
129. office friendships : *from, in*
130. oil crisis : *about*
131. oil well : *for*
132. olive oil : *from*
133. onion tears : *cause₂*
134. paper money : *make₂*
135. parent organization : *be, make₂*
136. party members : *have₂, in*
137. peanut butter : *make₂*
138. pedal extremities : *have₂*
139. people power : *from, have₂*
140. pet families : *have₁*
141. pet spray : *for*
142. pet theory : *be*
143. phantom limb : *be*
144. picture album : *for, have₁*
145. picture book : *have₁*
146. pine tree : *be*
147. plant food : *for*
148. plastic toys : *make₂*
149. plum wine : *from, make₂*
150. pole height : *have₂*
151. policy matters : *about*
152. pork suet : *from, have₂*
153. pot high : *cause₂*
154. pressure cooker : *use*
155. price dispute : *about*
156. queen bee : *be*
157. radio communication : *use*
158. rice paper : *from, make₂*
159. rye whiskey : *from, make₂*
160. salt lake : *have₁*
161. sand dune : *make₂*
162. sap tree : *have₁, make₁*
163. satellite nation : *be*
164. sea breeze : *from*
165. servant girl : *be*
166. sex scandal : *about*
167. shock treatment : *use*
168. sister node : *be*
169. smoke signals : *be, make₂, use*
170. snow blindness : *cause₂*
171. sob story : *cause₁*
172. soldier ant : *be*
173. song bird : *make₁*
174. sports activities : *be*
175. sports magazine : *about*
176. star shape : *be*
177. starvation diet : *use*
178. steam iron : *use*
179. steel helmet : *make₂*

- | | |
|---|---|
| 180. stone tools : <i>make</i> ₂ | 197. tire rim : <i>for</i> , <i>have</i> ₁ |
| 181. store clothes : <i>from</i> | 198. tobacco ash : <i>from</i> |
| 182. student committee : <i>make</i> ₂ | 199. vacuum cleaner : <i>use</i> |
| 183. student friends : <i>be</i> | 200. vapor lock : <i>cause</i> ₂ |
| 184. student power : <i>from</i> , <i>have</i> ₂ | 201. vegetable soup : <i>have</i> ₁ , <i>make</i> ₂ |
| 185. student problems : <i>about</i> , <i>have</i> ₂ | 202. voice vote : <i>use</i> |
| 186. sugar cube : <i>be</i> , <i>make</i> ₂ | 203. warrior caste : <i>make</i> ₂ |
| 187. summer months : <i>in</i> | 204. wastebasket category : <i>be</i> |
| 188. summer travels : <i>in</i> | 205. water drop : <i>be</i> , <i>make</i> ₂ |
| 189. surface tension : <i>in</i> | 206. water mark : <i>cause</i> ₂ |
| 190. suspense film : <i>cause</i> ₁ , <i>have</i> ₁ | 207. water wheel : <i>use</i> |
| 191. tape measure : <i>be</i> | 208. weekend boredom : <i>cause</i> ₂ , <i>in</i> |
| 192. target structure : <i>be</i> | 209. winter season : <i>be</i> |
| 193. tax law : <i>about</i> | 210. winter sports : <i>in</i> |
| 194. teaching profession : <i>about</i> , <i>be</i> | 211. women professors : <i>be</i> |
| 195. tear gas : <i>cause</i> ₁ | 212. wood shavings : <i>from</i> , <i>make</i> ₂ |
| 196. testtube baby : <i>from</i> | 213. worker teams : <i>have</i> ₁ , <i>make</i> ₂ |

ANNEXE C Vecteurs prédictifs de référence de Mturk fournis par Nakov

Vecteurs prédictifs pour toutes les réponses des formulaires

Abortion problem [ABOUT] involve(9); concern(6); be caused by(4); come from(3); surround(3); occur in(3); stem from(3); be related to(3); relate to(3); occur with(2); result from(2); deal with(2); affect(2); be about(2); include(1); arise during(1); be(1); be created by(1); mean(1); argue(1); originate from(1); effect(1); revolve around(1); be defined by(1); abhor(1); be found in(1); call for(1); discuss(1); threaten(1); function in(1); derive from(1); refer to(1); happen because of(1); happen with(1); arise in(1); lie with(1); complicate(1); be defined as(1)

Abortion vote [ABOUT] be about(7); concern(6); involve(6); allow(3); pertain to(3); support(3); affect(3); decide(2); decide on(2); approve(2); be against(2); regulate(2); prohibit(2); be for(2); refer to(2); address(2); determine(2); empower(2); legalize(1); be on(1); favor(1); protect(1); encourage(1); govern(1); promote(1); be taken regarding(1); be pro(1); feed(1); legislate(1); authorise(1); terminate in(1); decide about(1); add to(1); pass through(1); be in(1); come from(1); sanction(1); fight(1); define(1); reference(1); affirm(1); be regarding(1); exist with(1)

Accident weather [CAUSE₁] cause(21); promote(4); lead to(4); provoke(2); create(2); occur by(2); result in(2); contribute to(2); be conducive to(1); occasion(1); encourage(1); preclude(1); increase(1); influence(1); precipitate(1); bring on(1); happen by(1); induce(1); feed(1); generate(1); forbodes(1); be correlated with(1); be because of(1); percipitates(1); produce(1); trigger(1); be marked by(1); relate to(1); be associated with(1); be conducive to(1); accompany(1); foster(1); make(1); instigate(1); provide for(1)

Adolescent turmoil [IN] be experienced by(7); affect(7); occur in(6); involve(6); happen to(5); be caused by(3); surround(2); afflict(2); relate to(2); befall(2); bother(2); change(2); trouble(2); plague(2); come from(2); upset(2); effect(1); revolve around(1); foment inside(1); occur to(1); enrage(1); confuse(1); challenge(1); happen during(1); be felt by(1); traumatise(1); transpire in(1); attack(1); be connected with(1); emerge from(1); pertain to(1); embolden(1); depress(1); center around(1); be suffered by(1); stem from(1); be found in(1); torment(1); be exclusive to(1); manifest(1); be encountered by(1); appear in(1); deal with(1); be generated by(1); result from(1); traverse(1); pain(1); concern(1); be faced by(1)

Adventure story [ABOUT] contain(8); be about(7); involve(6); describe(6); include(4); tell of(4); have(4); relate(3); tell about(3); discuss(3); create(2); feature(2); revolve around(1); center around(1); center on(1); be of(1); narrate(1); be(1); recount(1); tell(1); incorporate(1); encompass(1); cover(1); explain(1); speak of(1); concern(1); be characterized by(1); exude(1); be found in(1); use(1); supply(1); provide(1); need(1); go into(1); highlight(1); be comprised of(1); be marked by(1); communicate(1); refer to(1); relate to(1); exhibit(1); concoct(1); explians(1); reference(1); imagine(1); talk about(1); require(1)

Air pressure [CAUSE₂] come from(9); be caused by(7); be due to(4); be made by(4); use(3); involve(3); be measured in(3); result from(2); be related to(2); pertain to(2); measure(2); occur in(1); force(1); be exerted by(1); consist of(1); effect(1); impact(1); be made from(1); emanate from(1); be cuased by(1); need(1); concentrate(1); emerge from(1); be in(1); be created by(1); derive from(1); be induced by(1); be developed from(1); relate to(1); be made up of(1); deal with(1); contain(1); utilize(1); be measured from(1); be overfilled with(1); focus(1); be exhibited by(1); be

filled with(1)

Alligator leather [FROM] come from(14); be made from(13); resemble(3); be taken from(3); be made of(3); look like(2); feel like(2); be(2); be created from(2); be found on(2); be from(2); be grown by(1); be on(1); belong to(1); be composed of(1); be skinned out of(1); imitate(1); be of(1); be removed from(1); use(1); be pulled from(1); be produced from(1); be made using(1); be derived of(1); start as(1); be sewn from(1); be made out of(1); be obtained from(1); contain(1); be supplied by(1); be sewed from(1); hide of(1); be fashioned from(1); cover(1); kill(1); consist of(1); be skin from(1); be born of(1)

Apple cake [HAVE₁] contain(16); be made from(10); be made with(9); be made of(7); have(6); taste like(6); include(4); use(3); come from(3); taste of(2); be baked with(2); be filled with(2); consist of(1); originate from(1); be topped by(1); be composed of(1); be flavored with(1); possess(1); start with(1); be prepared with(1); be made using(1); be made out of(1); smell of(1); depend on(1); utilize(1); employ(1); feature(1); be infused with(1)

Apple core [HAVE₂] come from(13); be inside(5); be in(4); be found in(3); be within(3); be located in(2); be from(2); fill(1); belong to(1); be derived from(1); spawn(1); be surrounded by(1); be of(1); reside in(1); reside within(1); be contained within(1); penetrate(1); create(1); exist in(1); remain of(1); emerge from(1); run through(1); support(1); be inside of(1); center(1); make up(1); reproduce(1); sit in(1)

Arms budget [FOR] pay for(7); be for(6); buy(5); concern(4); be meant for(3); include(3); purchase(2); be used for(2); deal with(2); cover(2); provide for(2); be spent on(2); be earmarked for(1); save for(1); be allocated to(1); be allocated for(1); supply(1); be linked to(1); provide(1); oversee(1); fund(1); be set aside for(1); list(1); be needed for(1); refer to(1); be connected to(1); account for(1); relate to(1); consider(1); be limited to(1); be saved for(1); be intended for(1); specify(1); manage(1); be utilized for(1); entail(1); affect(1)

Automobile plant [FOR] make(17); assemble(15); manufacture(15); build(11); produce(10); create(3); turn out(2); put together(2); construct(2); manufactures(1); put(1); outfit(1); fabricate(1); develop(1); knock out(1); churn out(1); be made of(1); test(1); crank out(1); reproduce(1)

Bacon grease [FROM] come from(22); be derived from(5); be rendered from(4); be made from(4); taste like(3); be produced by(3); be from(3); smell like(2); be contained in(2); be found in(2); come out of(2); emerge from(2); be in(2); be fried out of(1); drip from(1); cook out of(1); be used in(1); ooze from(1); be produced from(1); stick to(1); be exuded by(1); be drained from(1); be created by(1); be left by(1); derive from(1); happen with(1); drain from(1); be extracted from(1); render from(1); make(1); be cooked out of(1); result from(1); drip off of(1); drained from(1); separate from(1); coat(1); surround(1)

Basketball season [FOR] include(8); involve(5); feature(5); contain(4); have(3); be for(3); center around(2); promote(2); focus on(2); play(1); revolve around(1); be based around(1); be devoted to(1); occur during(1); precipitate(1); host(1); allow for(1); constains(1); be active for(1); offer(1); encompass(1); highlight(1); demonstrate(1); be comprised of(1); refer to(1); relate to(1); show-cases(1); come from(1); welcome(1); be best for(1); glorify(1); affect(1); concern(1); to play(1)

Bear country [HAVE₁] have(11); contain(9); be inhabited by(6); house(4); harbor(4); support(4); feed(3); protect(2); supply(2); sustain(2); be populated by(2); be populated with(2); nourish(2); shelter(2); belong to(1); be home to(1); be infested with(1); welcome to(1); possess(1); suggest(1); nurture(1); be conducive to(1); be known for(1); be lived in by(1); love(1); be habitated by(1); be native to(1); exist with(1)

Beehive hairdo [BE] look like(21); resemble(16); be shaped like(5); remind of(3); appear(2); emulate(2); be described as(1); be formed like(1); be referred to as(1); be modeled after(1); be modeled on(1); imitate(1); be like(1); be named after(1); appear like(1); rise like(1); copy(1); be styled like(1); duplicate(1)

Birth pains [CAUSE₂] be caused by(11); occur during(6); accompany(5); result from(4); be due to(4); precede(2); come from(2); coincide with(2); occur because of(2); develop during(2); happen during(2); be related to(2); be because of(2); be from(2); come before(1); announce(1); produce(1); come with(1); result in(1); prepare for(1); occur after(1); occur with(1); emit from(1); come during(1); be had during(1); emit during(1); involve(1); create(1); trigger(1); be induced by(1); happen with(1); occur before(1); happen at(1); occur at(1); complicate(1); anticipate birth(1); affect(1); come through(1)

Blanket excuse [BE] cover like(9); resemble(5); be like(5); act like(4); act as(2); work like(2); be used as(1); be analogous to(1); behave like(1); seem like(1); enclose like(1); function as(1); serve to(1); appear like(1); be given as(1); be(1); be described as(1); take care of like(1); include like(1); serve as(1); protect like(1); appear(1); emulate(1); envelop like(1); coat like(1)

Bronze statue [MAKE₂] be made of(20); be composed of(5); be(5); contain(5); be cast from(4); be made from(3); be sculpted in(2); be cast of(2); look like(2); be manufactured from(2); be created with(2); be made out of(2); be comprised of(2); be formed from(2); come from(2); cast in(1); be cast in(1); be in(1); be made with(1); be sculpted from(1); be colored(1); recall(1); be constructed out of(1); use(1); be constructed from(1); be molded from(1); emerge from(1); be built from(1); be fashioned from(1); appear(1); be designed in(1); be fabricated in(1); be forged in(1); consist of(1); emulate(1); be fabricated from(1)

Budget speech [ABOUT] be about(11); talk about(8); discuss(7); concern(5); involve(4); describe(4); refer to(4); address(3); explain(2); be on(1); stump for(1); revolve around(1); be made after(1); promote(1); focus on(1); be made before(1); clarify(1); elaborate on(1); announce(1); detail plan for(1); encompass(1); talk of(1); turn in(1); relate to(1); deal with(1); pertain to(1); come from(1); cover(1); lay out(1); present(1); reference(1); be made on(1); be made during(1); be concerned with(1)

Bull ring [FOR] contain(12); hold(4); be worn by(3); have(3); house(3); be for(3); resemble(2); exhibit(2); feature(2); surround(2); be used by(1); encompass(1); be in(1); look like(1); be on(1); constrain(1); restrict(1); include(1); corral(1); fence in(1); show off(1); involve(1); adorn(1); be placed upon(1); trap(1); appear on(1); demonstrate(1); be made for(1); be used for(1); keep(1); showcases(1); go through(1); be located in(1); be used on(1); exploit(1); be inhabited by(1)

Cable network [MAKE₂] consist of(5); be made of(5); be on(4); use(4); provide(3); run on(3); carry(2); be linked by(2); be watched on(2); be made up of(2); be connected by(1); have(1); be viewed on(1); deliver(1); run through(1); be shown on(1); be found on(1); be carried on(1); connect(1); utilize(1); require(1); rectify(1); resemble(1); involve(1); be aired on(1); broadcast use(1); be aired by(1); be broadcast on(1); be facilitated by(1); be joined by(1); connect use(1); work for(1); be supplied by(1); be carried by(1); be played on(1); broadcast over(1); come through(1); be relayed by(1); program(1); be transmitted by(1); be formed by(1); comprise of(1); control by(1); display on(1); encase(1); broadcast with(1); produce(1); enter(1); be featured on(1); be in(1); be located on(1); broadcasts(1); be built with(1); reach through(1); come from(1); be brought by(1); be formed of(1)

Cactus plant [BE] be(12); look like(9); be called(2); grow like(2); be related to(2); contain(2); be classified as(2); resemble(1); originate from(1); be composed of(1); be derived from(1); seem

like(1); form(1); be made of(1); grow into(1); create(1); be described as(1); be created by(1); be characterized as(1); be comprised of(1); derive from(1); relate to(1); bear(1); be categorized as(1); pertain to(1); appear(1); be like(1); consist of(1); be categorized with(1); be categorized as(1); birth(1); grow from(1)

Candy cigarette [MAKE₂] be made of(15); taste like(8); be made from(7); be(6); look like(5); be composed of(3); resemble(3); contain(3); create from(2); be eaten like(2); make from(1); be created out of(1); smell like(1); hide(1); use(1); be manufactured as(1); be formulated from(1); be created from(1); be sold as(1); be molded from(1); camouflage(1); emerge from(1); be formed in(1); be constituted by(1); appear as(1); make of(1); come from(1); appear(1); consist of(1); conceal(1); be made with(1); be fabricated from(1); taste of(1); disguise(1)

Cane sugar [FROM] come from(18); be made from(11); be derived from(5); be produced from(5); be created from(4); be processed from(4); be extracted from(3); be made of(3); originate from(2); be squeezed from(2); derive from(2); be found in(1); be from(1); be refined from(1); be grown in(1); be manufactured from(1); be taken from(1); be started from(1); be fabricated using(1); be scraped from(1); be made out of(1); emerge from(1); generate from(1); contain(1); be crushed from(1); consist of(1); be made with(1); be fabricated from(1); grow from(1)

Canine companion [BE] be(21); resemble(3); belong to(2); bark like(2); look like(2); be classified as(2); be from(2); be born(2); be made from(1); be of(1); howl like(1); act like(1); work as(1); put to work(1); eat like(1); like(1); be bred as(1); emerge from(1); derive from(1); pant like(1); come from(1); consist of(1); ennoble all(1); be trained as(1); exist as(1); assist as(1); descend from(1)

Cash basis [BE] involve(7); use(6); be based on(5); require(4); include(3); come from(3); depend on(3); consist of(2); be made of(2); accept(2); need(1); mean(1); utilize(1); demand(1); be based around(1); be formed by(1); be received in(1); be backed by(1); supply(1); begin with(1); be related to(1); be founded in(1); deal in(1); reflect in(1); emerge from(1); derive from(1); be founded on(1); result in(1); relate to(1); contain(1); be accounted in(1); be valued in(1); determine(1); rely on(1); be defined as(1); be from(1); be regarding(1); be concerned with(1); revolve from(1)

Ceiling price [BE] hit(5); be(5); reach(3); serve as(2); provide(2); form(2); go to(2); be at(2); stop at(2); approach(2); be used as(1); represent(1); be constrained by(1); constitute(1); define(1); maxes at(1); at(1); be terminated at(1); end at(1); describe(1); be for(1); be as high as(1); come from(1); be like(1); be capped by(1); seem(1); make(1); emulate(1); be maximized at(1)

Child actor [BE] be(22); look like(4); portray(3); start as(1); include(1); play(1); have(1); involve(1); act like(1); star as(1); work as(1); mimic(1); pass as(1); resemble(1); be classified as(1); substitute for(1); qualify as(1); act as(1)

Childhood dreams [IN] occur in(6); occur during(6); happen during(5); happen in(4); come from(3); originate in(3); come in(3); start in(2); relate to(2); enhance(2); be created during(2); be about(2); have(1); begin in(1); be(1); revolve around(1); belong to(1); manifest in(1); be had in(1); be dreamed(1); incubate during(1); involve(1); generate during(1); exist in(1); be related to(1); be during(1); be remembered in(1); be dreamt up in(1); invade(1); reference(1); infiltrate(1); transpire during(1); surround(1); be dreamt in(1); be created in(1); be spawned from(1); center around(1); be created at(1); reflect(1); be found in(1); come about in(1); be over after(1); be specific to(1); refer to(1); be remember from(1); accompany(1); be relegated to(1); result from(1); fulfill(1); color(1); be spawned in(1); happen(1)

Chocolate bar [MAKE₂] contain(17); be made of(16); be made from(10); taste like(7); be composed of(7); consist of(5); be(3); have(2); smell of(2); be manufactured from(2); be formed from(2); melt into(2); serve(1); sell(1); incorporate(1); be made with(1); be comprised of(1); be constituted

by(1); be solidified from(1); be flavored with(1); store(1); be flavoured with(1); be created from(1); taste of(1)

Cigarette burn [CAUSE₂] be caused by(21); come from(9); be made by(7); result from(6); be due to(4); originate from(2); be done with(2); emerge from(2); be left by(2); be created by(2); resemble(1); be inflicted by(1); be from(1); emit from(1); occur because of(1); be produced with(1); be given by(1); be blamed on(1); involve(1); ostracizes(1); be because of(1); be specific to(1); be induced by(1); happen because of(1); in regard to(1); be associated with(1); be started by(1); be originated from(1); be made with(1); be seared by(1)

Cigarette war [ABOUT] involve(12); concern(5); be about(5); be over(3); revolve around(2); be fought over(2); ban(2); focus on(2); be caused by(2); be caused for(1); include(1); be fought about(1); be against(1); discourage(1); occur over(1); happen for(1); target(1); arise from(1); eradicate(1); be motivated by(1); emerge from(1); control(1); be fought because of(1); denote(1); come from(1); be waged over(1); discount(1); differentiate(1); curtail(1); surround(1); oppose(1)

Citizen soldier [BE] be(20); represent(3); come from(3); act as(2); fight for(2); protect(1); defend(1); care for(1); start out as(1); guard(1); be selected from(1); live as(1); work as(1); relate to(1); be deemed(1); serve as(1); harbor(1); be verified as(1); support(1); be classified as(1); be defined as(1); emulate(1); be classified(1); be considered(1); concern(1)

City folk [IN] live in(21); reside in(13); come from(8); inhabit(5); stay in(2); occupy(2); be from(2); emerge from(2); be located in(2); be found in(2); thrive in(1); work in(1); dwell in(1); commute to(1); be used to(1); be accustomed to(1); prefer to live in(1); commute in(1); compose(1); enjoy(1); be urban like(1); be domiciled in(1); belong in(1)

City wall [HAVE₂] surround(24); protect(10); enclose(8); encircle(7); encompass(3); be in(3); contain(2); snake around(1); border(1); go around(1); isolate(1); guard(1); separate(1); be around(1); envelop(1); divide(1); defend(1); be built around(1); block(1); define(1); confine(1); fortify(1); belong to(1); be used by(1); contains(1); be outside of(1); surrounds(1); shield(1); border around(1); demarcate(1)

Coal dust [FROM] come from(20); be made from(5); be(4); be made of(4); be created by(4); derive from(4); be generated from(2); be composed of(2); emit from(2); be produced by(2); be produced from(2); emerge from(2); result from(2); be of(1); contain(1); accompany(1); originate from(1); be created out of(1); be attached to(1); go with(1); be found on(1); come with(1); look like(1); cover(1); be generated by(1); be spread by(1); consist of(1); be from(1); be shed from(1); come off(1)

Coffee nerves [CAUSE₂] be caused by(13); come from(8); result from(5); need(3); be affected by(2); be blamed on(2); emerge from(2); be from(2); be irritated by(1); occur because of(1); resemble(1); be brought on by(1); be aggravated by(1); be related to(1); be jangled from(1); be created from(1); be sourced from(1); be because of(1); be frayed from(1); be caused from(1); run on(1); emanate(1); be induced by(1); generate from(1); be stimulated by(1); be associated with(1); recognize(1); be generated by(1); descend from(1); be addicted to(1)

Coke machine [FOR] dispense(21); sell(19); contain(9); vend(7); distribute(6); hold(4); store(4); give(3); serve(2); supply(2); offer(2); give out(2); advertise(2); have(1); deliver(1); eject(1); market(1); administer(1); free(1); refrigerates(1); discharge(1); provide(1); meet out(1); expel(1); doles out(1); exchange for(1); dish out(1); display(1); chill(1); throw up(1)

College town [HAVE₁] have(15); contain(15); house(4); surround(4); host(3); support(3); be near(2); depend on(2); center around(1); include(1); be(1); rely on(1); be centered around(1); encou-

rage(1); be enthralled to(1); depend upon(1); be found near(1); border(1); center around(1); be attached to(1); supply(1); be linked to(1); involve(1); survive on(1); be sustained by(1); accommodate(1); encompass(1); own(1); accomodates(1); have within(1); nourish(1); benefit from(1); sustain(1); be focused on(1); be made up of(1); cater to(1); lie beside(1); be build near(1); be around(1); foster(1); harbor(1); begin because of(1); feature(1); result from(1); be inside(1); surround(1)

Collie dog [BE] be(12); look like(8); resemble(2); come from(2); belong to(2); be related to(2); be called(2); be classified as(2); be made from(1); be named(1); be implied to(1); indicate to(1); sire(1); start with(1); be bred(1); emerge from(1); generate from(1); be bred from(1); be categorized as(1); exhibit(1); woks like(1); harbor(1); be defined as(1); be from(1); be considered(1); be born(1); be born of(1)

Color television [HAVE₁] display(9); show(8); be in(7); display in(5); have(4); use(4); project in(3); broadcast in(3); transmit in(2); be found in(2); contain(2); show in(2); utilize(2); project(2); produce(1); play in(1); include(1); operate in(1); project with(1); screen with(1); show off(1); be of(1); supply(1); showcases(1); be broadcasted in(1); illuminate(1); employ(1); decode(1); receive(1); reproduce(1)

Communist tenet [IN] be believed by(10); be held by(6); be written by(5); support(4); belong to(3); be followed by(3); be made by(3); come from(3); emerge from(2); define(2); apply to(1); be defined by(1); cope with(1); break under(1); start with(1); regulate(1); hold for(1); be set down by(1); be adhered to by(1); deliver(1); provoke(1); demand of(1); talk to(1); relate to(1); be associated with(1); be promoted by(1); converse with(1); guide(1); make up(1); attract(1); explain(1); be exalted for(1); get to(1); originate from(1); labor over(1); issue from(1); excite(1); stem from(1); evolve from(1); be of(1); be sustainde for(1); agree with(1); be espoused by(1); be spoken by(1); purport(1); wonder about(1); be encompassed by(1); be believed in by(1); be created by(1); deal with(1); address(1); lead out(1); characterize(1); be practised by(1); be propositioned by(1); consist of(1); be upheld by(1); be professed by(1); affect(1); concern(1)

Company assets [HAVE₂] belong to(16); be owned by(14); be held by(3); be used by(3); be generated by(2); lie within(2); be possessed by(2); reside in(1); involve(1); be produced by(1); be under(1); be(1); be amassed by(1); benefit(1); be controlled by(1); be created by(1); derive from(1); sustain(1); enable(1); come from(1); help(1); support(1); strengthen(1); hold by(1); be purchased by(1); be employed by(1); affect(1); arise within(1); finance(1); fund(1); be exclusive to(1); enrich(1); underpin(1); pass through(1); happen within(1); be made by(1); be tied to(1); be earned by(1)

Concrete desert [MAKE₂] be made of(14); be made from(5); be composed of(5); resemble(5); consist of(5); contain(4); look like(3); be(2); be in(2); be covered in(1); include(1); feel like(1); be covered with(1); turn into(1); disappear under(1); seem like(1); be expressed in(1); appear like(1); be because of(1); be layered with(1); be for(1); harden as(1); be comprised of(1); be built from(1); be made up of(1); be turned in(1); be fashioned from(1); come from(1); embody(1); be caused by(1); utilize(1); be surfaced with(1); form from(1); be filled with(1)

Concussion force [CAUSE₁] cause(23); create(9); result in(5); produce(4); make(3); trigger(2); lead to(2); inflict(2); induce(2); form(2); bring about(2); result(1); relate to(1); pertain to(1); stimulate(1); be indicated by(1); be caused by(1); be found in(1); involve(1); generate(1); be produced by(1); instigate(1); force(1)

Cooking utensils [FOR] be used for(17); be used in(9); facilitate(4); help(3); aid(3); be required for(2); be found in(2); be utilized in(2); involve(2); function for(2); be needed for(2); enhance(2);

be used during(2); help in(1); be made for(1); be for(1); improve(1); be needed during(1); be meant for(1); be employed in(1); aid in(1); work for(1); partake in(1); enable(1); underscore(1); speed up(1); depend on(1); accompany(1); help with(1); work with(1); be intended for(1); be involved in(1); assist in(1); provide aid in(1); be utilized for(1); serve for(1); suit(1)

Copper coins [MAKE₂] be made of(12); be made from(10); contain(9); be composed of(5); consist of(4); be cast from(3); be made out of(3); include(2); be produced from(2); be comprised of(2); feature(2); have(1); be struck in(1); look like(1); be melted from(1); be derived from(1); be minted in(1); feel like(1); be stamped from(1); be manufactured from(1); be constructed from(1); be created from(1); be made using(1); be formed from(1); come from(1); utilize(1); be forged from(1); be formed of(1); be forged of(1)

Coriander curry [HAVE₁] contain(18); include(10); be made from(8); have(7); be made with(7); come from(4); taste like(3); be flavored with(3); use(3); be garnished with(2); smell of(2); feature(2); be mixed with(2); taste of(2); be composed of(1); destroy(1); be included in(1); be spiced with(1); remind of(1); exude(1); be make with(1); be(1); incorporate(1); be created from(1); be prepared with(1); dabble in(1); be made using(1); produce(1); deliver(1); emanate(1); burn with(1); look like(1); enhance with(1); utilize(1); consist of(1); evoke(1); be sweetened(1); enrich with(1); be cooked with(1); require(1); be about(1); be flavored by(1)

Country butter [FROM] be made in(18); come from(14); be produced in(6); be from(3); be associated with(3); originate in(2); be churned in(2); taste(1); be styled after(1); stand for(1); derive from(1); exemplify(1); be sold in(1); taste like(1); recall(1); remind of(1); make think of(1); be eaten in(1); be produced by(1); be developed in(1); be related to(1); come out of(1); taste of(1)

Country visitors [FROM] go to(9); come from(8); visit(8); travel to(7); live in(3); tour(3); hail from(3); be from(3); enjoy(2); explore(2); travel from(1); arrive from(1); represent(1); reside in(1); visit from(1); travel in(1); enter(1); be in(1); stay in(1); recreate in(1); journey to(1); holiday in(1); be denizens of(1); come to(1); emerge from(1); originate in(1); relax in(1); stay at(1); go in(1); stop at(1); flee(1); vacation in(1); runaway from(1); belong in(1)

Cream sauce [HAVE₁] contain(11); be made from(8); use(6); be made of(6); include(4); consist of(3); be composed of(3); involve(2); be like(2); resemble(2); be based on(2); taste like(1); have(1); incorporate(1); look like(1); taste of(1); be made with(1); require(1); be derived from(1); be shaped as(1); be of(1); seem like(1); contains(1); be smooth like(1); emerge from(1); have within(1); be made up of(1); result in(1); be stirred into(1); come from(1); be thickened by(1); rely on(1); feature(1); be thickened with(1); be feels like(1)

Daisy chains [MAKE₂] be made of(12); be made from(9); contain(5); consist of(5); be comprised of(4); look like(4); use(3); link(3); be created with(2); resemble(2); be created from(2); be built from(2); appear(2); be composed of(1); include(1); be connected like(1); cut(1); have(1); be constructed from(1); be shaped like(1); incorporate(1); showcase(1); be made using(1); interlock(1); be made up of(1); be woven of(1); connect(1); utilize(1); emulate(1); be made with(1); loop like(1)

Deficiency disease [CAUSE₂] be caused by(18); come from(6); result from(5); be due to(3); be associated with(2); cause(2); start with(2); be created by(2); happen because of(2); stem from(1); have(1); involve(1); create(1); arise from(1); result in(1); occur because of(1); be brought on by(1); begin with(1); be fed by(1); be related to(1); follow(1); begin by(1); occur due to(1); identify(1); yield to(1); be caused from(1); like(1); lead to(1); emerge from(1); control(1); derive from(1); generate from(1); refer to(1); relate to(1); pertain to(1); develop by(1); stem out of(1); ail(1); affect(1); bring about(1)

Desert rat [IN] live in(22); reside in(10); be found in(6); come from(6); inhabit(5); survive in(3); exist in(3); thrive in(2); feed in(2); dwell in(2); love(1); burrow beneath(1); evolve in(1); prefer(1); burrow within(1); occupy(1); emerge from(1); be comfortable in(1); run through(1); flourish(1); be adapted to(1); be located in(1); scurry in(1); survive(1); be native to(1); hide in(1); be from(1); populate(1); inhabitates(1)

Disaster flick [ABOUT] be about(12); involve(6); portray(6); describe(5); depict(3); show(3); concern(3); contain(2); be written about(1); revolve around(1); center around(1); include(1); center on(1); evolve from(1); simulate(1); reenact(1); discuss(1); be(1); revel in(1); be related to(1); incorporate(1); be created by(1); be based on(1); come from(1); relate(1); feature(1); tell about(1); recreate(1); talk about(1)

Disease germ [CAUSE₁] cause(20); spread(5); carry(4); create(4); produce(3); generate(3); start(2); promote(2); lead to(2); result in(2); propagate(1); bring about(1); influence(1); pass along(1); inflict(1); precipitate(1); be related to(1); result(1); initiate(1); manufacture(1); relate to(1); contain(1); come from(1); transfer(1); begin(1); instigate(1); pass(1); perpetuate(1)

Drug deaths [CAUSE₂] be caused by(19); result from(7); be due to(6); occur because of(5); involve(2); come from(2); occur from(2); be related to(2); occur due to(2); be induced by(2); happen because of(2); happen due to(2); include(1); be blamed on(1); happen from(1); be from(1); be influenced by(1); be affected by(1); emanate from(1); be precipitated by(1); bring on by(1); stem from(1); transpire because of(1); terminate with(1); start with(1); arise from(1); be attributed to(1); occur because(1); look like(1); be found with(1); be preceded by(1); imply(1); start from(1)

Enemy strength [HAVE₂] belong to(6); be possessed by(6); be owned by(3); be held by(2); be of(2); describe(2); be attributed to(2); be possessed by(2); evaluate(2); empower(2); be demonstrated by(1); count(1); arise out of(1); be preoccupied with(1); have(1); build up(1); be had by(1); be found in(1); occur within(1); lift(1); use(1); be linked to(1); involve(1); be made of(1); carry(1); acquired by(1); encompass(1); emerge from(1); be controlled by(1); relate to(1); pertain to(1); come from(1); characterize(1); be displayed by(1); illuminate(1); be caused by(1); be exerted by(1); be shown by(1); define(1); be exhibited by(1); be employed by(1); describes(1)

Energy emergency [ABOUT] involve(12); concern(6); relate to(4); be due to(3); be caused by(3); be about(3); require(2); emerge from(2); refer to(2); come from(2); be from(2); demand(1); affect(1); impact(1); be inspired by(1); be made from(1); restrict(1); include(1); emanate from(1); endanger(1); link to(1); need(1); focus on(1); be(1); threaten(1); overuse(1); comprise(1); ration(1); be created by(1); crunch(1); pertain to(1); be based on(1); problems with(1); determine(1); be created because of(1); tie up(1); result from(1); reduce(1); be regarding(1); be concerned with(1)

Evening hours [IN] occur in(13); be in(9); occur during(5); include(2); involve(2); happen during(2); be of(2); refer to(2); pertain to(2); happen in(2); start in(1); be during(1); be considered(1); fill(1); occur(1); constitute form(1); be included at(1); indicate(1); be contained in(1); pass in(1); be found in(1); happen by(1); transend into(1); be related to(1); comprise(1); encompass(1); suggest(1); fit into(1); come in(1); be categorized as(1); come at(1); happen at(1); come from(1); occur at(1); determine(1); fall during(1); be inside(1); make up(1); belong in(1)

Extension ladder [HAVE₁] have(16); provide(5); contain(4); use(4); include(3); supply(3); make(3); form(2); perform(2); utilize(2); be capable of(2); be used as(1); permit(1); allow(1); be augmented with(1); be able to have(1); be subjected to(1); lengthen by(1); pull out(1); add(1); lengthen use(1); generate(1); possess(1); incorporate(1); add on(1); offer(1); be in(1); give(1); lengthen through(1); extend(1); be made with(1); grow from(1); require(1); be able to make(1)

Faith cure [USE] rely on(6); come from(5); be caused by(5); require(4); depend on(3); rely upon(2); depend upon(2); occur because of(2); be due to(2); use(2); involve(2); begin with(2); be done by(2); be related to(2); happen because of(2); be brought about by(2); result from(2); be based on(1); confirm(1); happen through(1); occur through(1); be affected by(1); be made through(1); happen from(1); be enhanced by(1); happen by(1); need(1); be produced by(1); be rooted in(1); be except by(1); be given through(1); come from(1); be powered by(1); be because of(1); be ascribed to(1); come about because of(1); be induced by(1); relate to(1); happen due to(1); utilize(1); employ(1); arise because of(1); achieved by(1); work by(1); start from(1)

Family antiques [HAVE₂] belong to(17); be owned by(11); be passed down through(5); come from(4); be kept in(2); pass through(2); stay in(2); be kept by(2); be associated with(2); be in(1); be passed down in(1); remain with(1); be held by(1); be housed by(1); be linked to(1); inherit from(1); be related to(1); be collected by(1); be prized by(1); be handed down in(1); be connected with(1); descend in(1); be passed down within(1); relate to(1); be inherited from(1); be cherished by(1); be inherited by(1); be maintained by(1); be passed down by(1); be located in(1); be possessed by(1); be managed by(1); be discovered in(1); pass down(1); survive(1); be purchased from(1); concern(1)

Family problems [IN] involve(5); affect(5); occur within(3); plague(3); happen in(3); be experienced by(2); occur in(2); be found in(2); be solved by(2); arise in(2); be caused by(2); concern(2); belong to(2); effect(1); destroy(1); include(1); exist within(1); be suffered by(1); hurt(1); challenge(1); trouble(1); run in(1); use(1); crop up in(1); appear in(1); be created by(1); disturb(1); be within(1); originate in(1); deal with(1); happen between(1); remain in(1); be generated in(1); be had by(1); mess up(1); be related to(1); happen within(1); originate with(1); be worked on by(1); happen to(1); complicate(1); break up(1)

Farm boy [FROM] work on(23); live on(22); come from(9); be born on(3); reside on(3); work in(3); grow up on(2); be raised on(2); toil on(2); labor on(2); reside at(2); live near(1); work(1); help on(1); love(1); originate from(1); be employed at(1); identify with(1); own(1); like(1); toil at(1); be associated with(1); be like(1); enjoy(1); reside near(1); be employed by(1); thwart(1)

Fatigue headache [CAUSE₂] be caused by(19); result from(9); come from(7); be due to(4); be brought on by(4); be produced by(3); happen because of(3); indicate(2); be related to(2); follow(2); emerge from(2); be created by(2); arise from(1); be induced by(1); coincide with(1); originate from(1); occur because of(1); occur do to(1); show(1); bring on by(1); stem from(1); be blamed on(1); be linked to(1); feed(1); develop from(1); involve(1); come from(1); correspond to(1); happen after(1); be associated with(1); be triggered by(1); be complicated by(1); cause(1); warn of(1)

Field mouse [IN] live in(22); come from(9); be found in(7); inhabit(7); reside in(4); dwell in(3); breed in(2); habitates(2); hide in(2); burrow in(1); run in(1); be in(1); travel through(1); live on(1); frolic in(1); procreate in(1); roam around(1); occupy(1); wander around(1); enter(1); emerge from(1); run through(1); feed in(1); generate from(1); originate in(1); reside(1); be native to(1); populate(1); be from(1); reside near(1)

Financing dilemma [ABOUT] involve(9); be caused by(6); concern(6); be related to(4); come from(4); affect(4); relate to(3); result from(3); be about(3); be due to(2); emerge from(2); complicate(2); cause(2); occur with(1); muddle(1); prevent(1); occur in(1); include(1); depend upon(1); be posed in(1); drain(1); cut into(1); occur during(1); inflict(1); be found in(1); be linked to(1); be(1); deficits(1); worry about(1); be in(1); be led by(1); happen to(1); regard(1); lose(1); problematizes(1)

Finger cymbals [USE] be played with(8); be worn on(8); be played by(5); be used by(4); use(3);

attach to(3); go on(2); resemble(2); need(2); be placed on(2); look like(2); require(2); be on(1); be put(1); be put on(1); be activated by(1); be clapped by(1); be held by(1); work on(1); operate by(1); wear on(1); be percussed by(1); be attached to(1); involve(1); be struck by(1); be(1); be played using(1); be found on(1); be sized for(1); emerge from(1); be tied to(1); fit on(1); be strapped on(1); appear(1); be waved on(1); be used with(1); be used on(1); clink on(1); be situated on(1); play by(1); be hit by(1); be designed for(1)

Finger lakes [BE] resemble(15); look like(12); be shaped like(7); emerge from(2); be thin like(2); flow in(2); remind of(2); appear like(2); have(1); contain(1); appear(1); be like(1); be narrow like(1); seem(1); consist of(1); emulate(1); mirror(1)

Fish scales [HAVE₂] cover(11); protect(8); be on(7); weigh(6); be found on(6); grow on(6); come from(5); appear on(2); coat(2); shed from(1); fall from(1); be grown by(1); insulate(1); be worn by(1); be all over(1); armour(1); adorn(1); be in(1); remove from(1); typify(1); derive of(1); mark(1); be grown on(1); enclose(1); shield(1); be layered on(1); grow from(1); overlap(1)

Flounder fish [BE] be(9); look like(6); resemble(3); be called(3); be classified as(3); taste like(2); be identified as(2); be named(1); smell like(1); group with(1); perform(1); struggle like(1); be trapped to(1); desire to(1); be related to(1); do(1); be categorized as(1); be cursed to(1); be destined to(1); be sold as(1); resembles(1); want to(1); be categorized as(1); come from(1); swim like(1); appear(1); be like(1); be seen in(1); seem(1); be doomed to(1); love to(1); be known to(1); be said to(1); seem to(1)

Flu virus [CAUSE₁] cause(19); spread(4); give(4); result in(3); create(3); infect with(3); contain(3); be(2); carry(2); induce(1); produce(1); look like(1); make(1); incubate into(1); exacerbate(1); turn into(1); happen from(1); transmit(1); be made of(1); involve(1); generate(1); breed(1); be related to(1); sicken with(1); lead to(1); intensify be(1); disseminate(1); come from(1); be implicated in(1); appear(1); instigate(1); be conceived by(1); bring about(1)

Frog man [BE] resemble(11); swim like(9); look like(6); act like(4); be like(3); be(2); dive like(2); emerge from(2); sell(1); study(1); love(1); model(1); behave like(1); function like(1); live with(1); dress like(1); imitate(1); seem like(1); swim with(1); appear like(1); impersonate(1); descend in(1); capture(1); sound like(1); move like(1); come from(1); appear(1); work with(1); maneuver like(1); cook(1); emulate(1); work like(1); be from(1); descend from(1)

Fruit tree [HAVE₁] bear(20); produce(16); grow(15); have(6); give(4); provide(3); develop(2); supply(2); make(2); hold(1); contain(1); bare(1); be laden with(1); be grown for(1); be filled with(1); be made from(1); bloom(1); issue(1); blossom with(1); create(1); be raised for(1); be covered by(1)

Future shock [CAUSE₂] occur in(6); involve(5); happen in(5); come from(4); be caused by(4); concern(3); be about(3); happen during(2); happen because of(2); result from(2); predict(2); come in(1); revolve around(1); emerge because of(1); emit from(1); occur because of(1); remain(1); occur during(1); be due to(1); be in reference to(1); occur within(1); be found in(1); prophecies(1); stream from(1); be linked to(1); foretell(1); be confused by(1); be related to(1); originate with(1); flow into(1); be relative to(1); refer to(1); relate to(1); take in(1); begin(1); occur from(1); be dismayed about(1); happen(1); be from(1)

Garter snake [BE] look like(13); resemble(9); be called(2); have(2); appear(2); be named after(2); eat(1); like(1); take away(1); be thin like(1); be named for(1); be named(1); be banded like(1); be classified as(1); consist of(1); call(1); be made of(1); wear(1); crawl through(1); reproduce(1)

Gas stove [USE] use(20); run on(9); burn(8); cook with(6); utilize(4); emit(3); be heated by(2); need(2); consume(2); work with(2); require(2); heat with(2); heat by(2); run by(2); employ(2);

leak(1); work on(1); run off(1); ignite(1); be fueled by(1); rely on(1); bake with(1); boil with(1); combust(1); fly with(1); operate with(1); be run by(1); saute with(1); be powered by(1); give off(1); exumes(1); be controlled by(1); contain(1); be based on(1); cook use(1); light(1); heat use(1); transform(1)

Glass eye [MAKE₂] be made of(19); be composed of(11); be made from(6); be(4); be created from(3); consist of(3); be constructed of(2); resemble(2); look like(2); be like(2); be created out of(2); be manufactured from(2); contain(2); come from(2); be of(1); be replaced by(1); originate from(1); be formulated of(1); imitate(1); be constructed out of(1); be produced in(1); be constructed from(1); incorporate(1); be produced from(1); emerge from(1); be in(1); be modelled from(1); appear(1); be worn with(1); be originated from(1); be fabricated from(1)

Government employment [IN] involve(4); be provided by(4); be sponsored by(3); be with(3); benefit(3); be made by(2); be within(2); come from(2); be offered by(2); be run by(1); assist(1); support(1); belong to(1); be initiated by(1); serve(1); occur for(1); be by(1); be supported by(1); be given by(1); stem from(1); be issued by(1); feed(1); happen for(1); begin with(1); be paid for by(1); service(1); be controlled by(1); derive from(1); be funded by(1); pertain to(1); be granted by(1); help(1); honor(1); happen in(1); be terminated by(1); be performed by(1)

Government land [HAVE₂] be owned by(23); belong to(15); be managed by(4); be controlled by(3); be possessed by(3); be administered by(2); come from(2); be maintained by(2); be occupied by(1); be regulated by(1); be seized by(1); be developed by(1); be used by(1); supply(1); be sold by(1); be bought by(1); descend in(1); be set aside by(1); be of(1); be run by(1)

Grain alcohol [FROM] be made from(16); come from(10); be produced from(7); be distilled from(5); contain(5); use(3); originate from(2); be made of(2); be fermented from(2); be derived from(1); be manufactured from(1); require(1); be found in(1); involve(1); incorporate(1); be created from(1); be processed from(1); be made out of(1); emulate from(1); emerge from(1); be crafted from(1); be extracted from(1); consume(1); be manufactured from(1); be generated by(1); rely on(1); consist of(1); transform(1); be made with(1)

Growth hormone [CAUSE₁] promote(11); cause(9); stimulate(6); affect(4); encourage(3); help(3); increase(2); regulate(2); enhance(2); produce(2); enable(2); allow(2); influence(1); induce(1); create(1); help in(1); control(1); accelerate(1); result in(1); fuel(1); help with(1); support(1); speed(1); effect(1); supply(1); generate(1); be necessary for(1); be related to(1); spark(1); supplement(1); quicken(1); implement(1); lead to(1); nourish(1); facilitates(1); aid(1)

Gutter language [FROM] come from(9); originate from(5); be used in(5); be from(5); resemble(4); belong in(3); be found in(2); stem from(2); emerge from(2); sound like(2); be derived from(1); begin in(1); be heard in(1); represent(1); be worthy of(1); remind of(1); be born in(1); involve(1); stink like(1); lower to(1); be in(1); derive from(1); relate to(1); pertain to(1); be associated with(1); appear from(1); be spoken in(1)

Hairpin turn [BE] look like(16); resemble(14); be shaped like(5); mimic(4); be curved like(2); be abrupt like(1); turn as(1); ensnare like(1); be like(1); be shaped as(1); remind of(1); be symbolized by(1); turn like(1); use(1); be sharp like(1); wind in(1); be curvy like(1); be described as(1); follow(1); flow like(1); twist in(1); appear(1); zigzag like(1); be like(1); twist like(1); consist of(1); emulate(1)

Hand brake [USE] be operated by(11); be used by(7); require(6); use(5); be activated by(5); be made for(4); be applied by(3); operate by(3); be controlled by(3); need(3); be worked by(2); stop by(2); work by(2); be handled by(2); utilize(1); be pulled by(1); be within reach of(1); be by(1); be operated with(1); be manipulated with(1); be pulled with(1); be run by(1); be available to(1);

involve(1); be moved with(1); be powered by(1); be applied by(1); take(1); be activated via(1); be used with(1); go in(1); stay inside(1); stop with(1); be pushed with(1); be engaged by(1)

Handlebar mustache [BE] look like(20); resemble(17); be shaped like(6); appear(2); appear like(2); be used as(1); clone(1); be formed like(1); suggest(1); mimic(1); curl like(1); recall(1); remind of(1); simulate(1); come from(1); be contoured like(1); be like(1); form(1); be akin to(1); spread out like(1); be close to(1)

Head noun [BE] be(6); describe(3); involve(2); be found in(2); refer to(2); be at(2); come at(2); stand at(1); have(1); form(1); act as(1); be called(1); become(1); occur at(1); include(1); represent(1); challenge(1); be located at(1); head(1); act like(1); lead by(1); emerge from(1); be referred to as(1); jump to(1); come from(1); be found at(1); be classified as(1); illustrate(1); make(1); act on(1); touch(1); start at(1); be placed at(1); begin at(1); lead(1); affect(1); be about(1); perform as(1)

Headache pills [FOR] cure(14); relieve(8); reduce(5); help(4); treat(4); ease(3); do away with(2); decrease(2); alleviate(2); heal(2); soothe(2); improve(2); prevent(1); suppress(1); be for(1); eliminate(1); stop(1); help with(1); deaden(1); drain(1); be taken during(1); negate(1); subdue(1); curb(1); knock out(1); diminish(1); feed(1); suppress(1); attack(1); control(1); quell(1); be used for(1); mitigate(1); solve(1); fix(1); eradicate(1); act upon(1); heel(1); be taken for(1); aid(1); minimize(1); lessen(1); curtail(1); ameliorate(1); affect(1); palliate(1)

Heart design [BE] look like(10); resemble(8); be shaped like(8); contain(6); include(5); use(3); incorporate(3); involve(3); feature(2); be made from(2); be shaped as(2); be made of(2); have(1); form(1); nourish(1); be made up of(1); come from(1); warm(1); be drawn like(1); be composed of(1); center around(1); involves(1); be created out of(1); promote(1); imitate(1); show off(1); be made to resemble(1); be engineered from(1); be related to(1); exist of(1); encompass(1); be made out of(1); emerge from(1); be comprised of(1); relate to(1); be shaped in(1); expose(1); appear(1); display(1); symbolize(1); consist of(1)

Heat rash [CAUSE₂] be caused by(26); come from(12); result from(5); be created by(4); stem from(3); originate from(2); be instigated by(1); be associated with(1); be exacerbated by(1); bespeak(1); be sensitive to(1); be derived from(1); react to(1); be due to(1); indicate to(1); be brought on by(1); be related to(1); come from(1); worsen in(1); be produced from(1); produce(1); come on from(1); emerge from(1); be induced by(1); be generated from(1); incubate in(1); be started by(1); be triggered by(1); occur from(1); be caused by(1); be from(1); suffer under(1)

Hermit crab [BE] act like(10); behave like(7); resemble(7); be(6); live like(6); hide like(2); look like(2); emulate(2); stand for(1); isolate like(1); model(1); imitate(1); remind of(1); recall(1); copy(1); live as(1); behave live(1); emerge from(1); mimic(1); parody(1); be called(1); exist like(1); come from(1); move like(1); be recluse like(1); attract(1); be isolated like(1)

History conference [ABOUT] discuss(11); be about(8); focus on(5); involve(4); concern(4); deal with(3); talk about(3); be on(2); center on(2); concentrate on(2); be related to(2); teach(2); occur in(1); explore(1); look at(1); examine(1); study(1); address(1); present(1); revolve around(1); debate(1); scrutinize(1); encourage(1); discover(1); inform about(1); profile(1); detail(1); highlight(1); elucidate(1); be steeped in(1); refer to(1); relate to(1); pertain to(1); speak about(1); analyze(1); delve into(1); evaluate(1); determine(1); confer about(1); handle about(1); reference(1); extrapolate on(1); share(1); explain(1); be concerned with(1); descend from(1)

Home remedy [FROM] be made at(12); come from(11); be used at(6); be used in(4); be found at(3); originate at(3); originate from(2); happen at(2); be developed at(2); be prepared at(1); be practiced in(1); be found in(1); be done at(1); be administered at(1); be prepared in(1); be

provided at(1); be created at(1); be discovered at(1); locate in(1); be designed at(1); help at(1); originate in(1); be limited to(1); be made in(1); use be(1); start at(1); heal at(1)

Honey bee [MAKE₁] make(25); produce(16); create(9); manufacture(5); store(4); eat(2); provide(2); secrete(2); generate(1); gather(1); formulate(1); manufactures(1); cultivate(1); turn out(1); synthesize(1)

Horse doctor [FOR] treat(19); heal(9); care for(6); cure(5); tend to(4); medicate(4); work with(4); work on(3); fix(3); help(3); serve(2); examine(2); mend(1); diagnose(1); operate on(1); empathize with(1); practise on(1); practice on(1); mdeicates(1); dress(1); deal with(1); specialize in(1); check on(1); attend to(1); aid(1)

Hydrogen bomb [USE] contain(9); use(8); be made from(5); be composed of(5); be made of(5); fuse(3); utilize(3); consist of(3); have(2); involve(2); need(2); be comprised of(2); ignite(1); split(1); explode because of(1); include(1); react to(1); waste(1); have in(1); break down(1); be created from(1); incorporate(1); degrade(1); disperse(1); be manufactured with(1); be made up of(1); be fueled by(1); consume(1); explode(1); come from(1); release(1); depend on(1); explode use(1); work with(1); scatter(1); employ(1); transform(1); exploit(1); require(1)

Immigrant minority [MAKE₂] be comprised of(9); include(7); be composed of(6); consist of(6); be made up of(6); contain(6); be made of(5); be(2); represent(2); involve(2); come from(2); descend from(1); belong to(1); grow by(1); speak for(1); have(1); be inclusive of(1); form(1); be populated with(1); begin with(1); encompass(1); be enlarged by(1); relate to(1); enumerate(1); be labeled as(1); form from(1)

Infant colonies [BE] contain(7); be made of(4); be composed of(3); resemble(3); be like(3); be(2); be comprised of(2); be made up of(2); support(2); consist of(2); have(1); behave like(1); include(1); be populated by(1); limit(1); remind of(1); be still(1); involve(1); house(1); incorporate(1); compose of(1); take after(1); sustain(1); be occupied by(1); shelter(1); be compose of(1); be compared to(1); begin(1); be lived in by(1); cordon off(1)

Job tension [CAUSE₂] be caused by(14); come from(8); result from(6); be related to(5); be created by(5); be due to(3); involve(3); be associated with(2); be found in(2); emerge from(2); be in(2); be carried over from(1); occur in(1); occur because of(1); occur during(1); be instigated by(1); emanate from(1); interrupt for(1); be blamed on(1); infect(1); run in(1); be evoked by(1); begin with(1); look at(1); start with(1); occur on(1); strike with(1); happen because of(1); arise in(1); occur at(1); be generated by(1); be located in(1); build from(1); invade(1); be triggered by(1); sour(1); occur from(1); worsen on(1); arise at(1); fester from(1); be from(1)

Juvenile court [FOR] try(13); deal with(8); judge(5); handle(4); adjudicate(3); prosecute(3); process(3); be for(3); involve(2); work with(2); sentence(2); protect(2); be concerned with(2); serve(1); help(1); listen to(1); adjucates for(1); be meant for(1); govern(1); punish(1); convict(1); rule on(1); discuss(1); focus on(1); be created for(1); exonerate(1); service(1); be made for(1); relate to(1); contain(1); cater to(1); pertain to(1); be associated with(1); specialize in(1); discipline(1); monitor(1)

Kennel puppies [FROM] live in(17); come from(9); be born in(8); be raised in(6); be bred in(6); be found in(3); be from(3); be adopted from(2); be housed in(2); inhabit(2); be lived in(2); be bred by(1); be kept in(1); grow up in(1); belong to(1); be trained in(1); be sold from(1); reside in(1); start in(1); be sold by(1); be keep in(1); emerge from(1); generate from(1); be boarded at(1); stay in(1); originate in(1); be kept(1); be located in(1); hail from(1); be protected inside(1); initiate in(1); start from(1); descend from(1)

Lace handkerchief [HAVE₁] be made of(18); be made from(7); be composed of(4); contain(4); consist of(4); be trimmed with(3); be(3); be decorated with(3); resemble(3); use(3); be comprised of(3); look like(3); be made with(3); have(2); be constructed from(2); be trimmed in(2); feature(1); be derived from(1); be bordered by(1); be cut from(1); be sewn with(1); be constructed of(1); be of(1); be patterned after(1); be fabricated of(1); be created from(1); be made out of(1); be designed from(1); showcases(1); be sewed from(1); come from(1); utilize(1); employ(1); be woven from(1); boast of(1); be constructed with(1); be edged with(1)

Language riots [ABOUT] be caused by(8); involve(6); be about(5); include(3); concern(3); be inspired by(2); happen because of(2); come from(2); result from(2); occur with(1); be caused for(1); demand(1); be fomented by(1); occur because of(1); be formed by(1); stem from(1); use(1); be rooted in(1); be related to(1); be because of(1); deploy(1); be incited because of(1); be comprised of(1); be created by(1); generate from(1); be induced by(1); be provoked by(1); be connected to(1); relate to(1); deal with(1); pertain to(1); be based on(1); happen due to(1); utilize(1); be brought about by(1); arise because of(1); employ(1); be classified as(1); be contributed to by(1); be commenced by(1); be regarding(1)

Laugh wrinkles [CAUSE₂] be caused by(17); come from(15); be created by(4); result from(4); bespeak(1); be derived from(1); show(1); emanate from(1); be caused by(1); stem from(1); be due to(1); be deepened by(1); be brought on by(1); be gotten from(1); develop from(1); appear with(1); be related to(1); follow(1); be produced from(1); show from(1); emerge from(1); be made by(1); generate from(1); appear as(1); happen because of(1); be associated with(1); appear during(1); show as(1); occur from(1); form from(1)

Lemon peel [HAVE₂] come from(19); surround(6); protect(5); cover(5); be found on(4); smell like(4); come off(4); be taken from(3); be made from(2); be from(2); grow on(1); be on(1); be derived from(1); emerge from(1); be located on(1); taste like(1); come off of(1); be extracted from(1); be removed from(1); envelop(1); adhere to(1); be outside of(1)

Lightning rod [FOR] attract(19); protect from(7); conduct(6); channel(5); draw(3); deflect(3); direct(3); absorb(2); divert(2); dissipate(2); arrest(2); draw off(1); be attractive to(1); be targeted by(1); defend from(1); invite(1); nullifies(1); redirect(1); be related to(1); trap(1); route(1); keep from(1); get hit by(1); save from(1); send(1); safeguard from(1); be used for(1); stop(1); contain(1); help with(1); insure against(1); focus(1); manage(1); be hit by(1); ameliorate(1); take care of(1); centralize(1)

Lion cub [BE] be(9); come from(9); grow into(6); become(5); be born of(4); be birthed by(3); be born from(3); be sired by(2); bear of(2); belong to(2); resemble(2); emerge from(2); descend from(2); be suckled by(1); mature into(1); transform into(1); turn into(1); be included in(1); emerge(1); be fed by(1); be borne of(1); propagaes(1); live among(1); be created by(1); look like(1); be protected by(1); roar like(1); be classified as(1); be parented by(1); be raised by(1); be from(1); be born(1); parent by(1)

Love song [ABOUT] be about(17); describe(6); talk about(5); concern(4); involve(3); speak of(3); refer to(3); demonstrate(2); deal with(2); mention(2); praise(2); express(2); discuss(2); ruminate on(1); celebrate(1); be written about(1); be inspired by(1); show(1); profess(1); remind of(1); mourn(1); romanticize(1); address(1); contain(1); speak about(1); sing of(1); sing about(1); encourage(1); vocalize(1); talk of(1); nourish(1); exemplify(1); croon of(1); come from(1); be caused by(1); pour out(1); reference(1); glorify(1); imply(1); teach about(1)

Machine translation [USE] be done by(13); be performed by(6); come from(6); use(3); be produced by(3); be developed by(3); be completed by(3); be made by(3); require(3); originate from(2); in-

volve(2); be created by(2); be generated by(1); be provided by(1); be processed by(1); be delivered by(1); occur in(1); issue from(1); be programmed into(1); be by(1); emanate from(1); be conducted by(1); be specific for(1); complete by(1); be done with(1); sprout from(1); be accomplished by(1); incorporate(1); service(1); spring from(1); generate from(1); refer to(1); be altered in(1); defecate from(1); pertain to(1); be calculated by(1); utilize(1); be meaningful for(1); be altered by(1); be from(1); be effected by(1); explain(1); concern(1)

Marine life [IN] be(9); exist in(4); live in(3); be enjoyed by(2); dwell in(2); originate from(2); be lived by(2); be of(2); be found in(2); be described as(2); come from(2); support(1); occur in(1); encourage(1); drain(1); be supported by(1); live(1); feed(1); supply(1); involve(1); be related to(1); baby(1); camouflage(1); flow into(1); turn(1); be called(1); send off(1); sober up(1); put down(1); live by(1); be exemplified by(1); be defined as(1); be given from(1); descend from(1); belong in(1); terrify(1); emit from(1); remain(1); train up(1); emanate from(1); reside in(1); show off(1); reward(1); occupy(1); smile at(1); deploy(1); wonder about(1); surprise(1); inhabit(1); refer to(1); berate(1); be categorized as(1); be led by(1); bring out(1); characterize(1); survive in(1); keep up(1); be lived in(1); be thought of as(1); show up(1); found in(1)

Midnight snack [IN] be eaten at(19); be consumed at(11); happen at(7); occur at(5); occur around(2); be eaten around(2); be made at(2); be snuck at(2); be taken at(2); be consumed around(1); appeal at(1); come by(1); be prepared at(1); be enjoyed at(1); come around(1); be served at(1); be had at(1); be eaten near(1); occur near(1); go into(1); satisfy at(1); be desired at(1); be consumed near(1); be sneaked at(1); happen near(1); enter(1); enjoy at(1); happen around(1); be craved at(1); pertain to(1); be at(1); come at(1); accompany(1); be devoured at(1); around(1); eat in(1); nourish at(1); come after(1); happen(1); gratify at(1)

Milieu therapy [USE] use(8); involve(4); utilize(4); modify(2); be based on(2); believe in(2); proceed in(1); occur in(1); be based upon(1); correct(1); change(1); be applied to(1); hope for(1); host(1); be performed in(1); work on(1); be(1); be held in(1); repair(1); describe(1); be accomplished in(1); be in(1); be used for(1); be induced by(1); treat(1); consider(1); point to(1); be caused by(1); be changed by(1); apply(1); work with(1); employ(1); practice(1); invoke(1); socialize among(1); be used on(1); happen in(1); require(1); concern(1)

Mining engineer [FOR] work in(9); study(9); specialize in(6); know about(6); do(3); work with(3); design(3); understand(2); know(2); investigate(2); learn about(2); focus on(2); deal with(2); be concerned with(2); build(1); specialise in(1); be qualified for(1); research(1); center on(1); work at(1); work on(1); be employed in(1); be occupied with(1); discuss(1); involve(1); contemplate(1); examine(1); think about(1); relate to(1); handle(1); apply(1); practice(1); be involved in(1); be employed by(1); talk about(1); improve(1); be degreed in(1)

Morning prayers [IN] be said in(12); be recited in(4); occur in(4); start(3); be done in(2); be said(2); be said at(2); be recited at(2); relate to(2); occur at(2); be made in(2); be written for(2); happen in(2); be spoken in(2); be recited(1); begin in(1); be performed in(1); be said during(1); be in(1); be chanted in(1); be recited during(1); be meant for(1); be observed in(1); start off(1); occur during(1); happen during(1); be spoken at(1); be prayed in(1); terminate in(1); be performed at(1); be uttered at(1); be offered up in(1); arise in(1); enervate(1); happen at(1); energize(1); be remembered in(1); be uttered in(1); be made during(1); be given during(1); belong in(1)

Morphology lecture [ABOUT] be about(13); explain(6); discuss(5); describe(5); concern(4); involve(3); teach(3); cover(3); include(2); deal with(2); teach about(2); debate(1); realize(1); attribute(1); be taught about(1); go in depth into(1); focus on(1); elaborate on(1); encompass(1); detail(1); highlight(1); study(1); address(1); pertain to(1); be based on(1); outline benefit of(1);

enlighten on(1); feature(1); define(1); be regarding(1); talk about(1)

Moth hole [CAUSE₂] be made by(16); be created by(8); be caused by(6); be eaten by(5); be produced by(4); come from(3); be generated by(2); result from(2); be perpetrated by(1); protect(1); emanate from(1); be blamed on(1); be used by(1); feed(1); be done by(1); be dug by(1); be because of(1); be made for(1); suggest(1); be left by(1); eat by(1); signify(1); conceal(1); be evidence of(1); be chewed by(1); be eaten out by(1)

Mother church [BE] act as(8); act like(6); be(3); be like(3); resemble(2); serve as(2); feel like(2); behave as(2); emulate(2); be supported by(1); play(1); support(1); be spawned by(1); be originated with(1); treat like(1); remind of(1); appeal to(1); serve to(1); recruit(1); be treated as(1); be created by(1); designate as(1); originates from(1); branch off from(1); appear(1); consist of(1); operate as(1)

Mountain lodge [IN] be in(9); be located in(9); be built in(6); be on(5); be found in(5); be situated on(4); be built on(2); be found on(2); be near(2); be nestled in(1); be surrounded by(1); top of(1); rest on(1); be situated in(1); be up in(1); be atop(1); lie in(1); view(1); be build on(1); be located on(1); be amidst(1); be around(1); stand on(1); run along(1); build in(1); be up(1); found in(1); sit in(1)

Mountain range [MAKE₂] be made of(12); contain(8); be composed of(5); have(5); include(3); be made up of(3); be of(2); comprise(2); consist of(2); be comprised of(2); hold(1); be(1); encompass(1); find(1); involve(1); be created from(1); be created by(1); be made by(1); be constituted by(1); run through(1); relate to(1); be formed from(1); display(1); feature(1); be formed of(1); comprises(1); be filled with(1)

Movement schisms [IN] divide(5); involve(4); create(2); destroy(2); cause(2); shatter(2); come from(2); affect(2); split(1); occur with(1); occur to(1); occur in(1); be characterized by(1); inflict(1); be due to(1); have(1); discourage(1); be produced by(1); be related to(1); be created from(1); catastrophize(1); disrupt(1); emerge from(1); result in(1); happen to(1); be located in(1); be caused by(1); break up(1); split from(1); happen in(1); be from(1); include(1); emanate from(1); be noted by(1); develop in(1); be found in(1); use(1); possess(1); force(1); be created by(1); be in(1); form within(1); separate(1); utilize(1); result from(1); instigate(1); concern(1); require(1); within(1)

Murder charge [BE] involve(14); accuse of(4); come from(3); concern(3); accuse(2); emerge from(2); relate to(2); consist of(2); imply(2); include(1); be for(1); revolve around(1); presume(1); emanate from(1); indicate(1); be due to(1); be of(1); come because of(1); declare(1); be related to(1); charge(1); indict for(1); be made for(1); blame(1); suggest(1); be labeled(1); refer to(1); be brought up for(1); mean(1); assume(1); claim(1); point to(1); determine(1); be generated by(1); specify(1); result from(1); implicate(1); presuppose(1); warn of(1); allege(1); be concerned with(1); be about(1)

Music box [MAKE₁] play(19); make(12); produce(10); emit(5); create(4); contain(4); provide(2); generate(2); give off(2); include(1); have(1); supply(1); emanate(1); amplify(1); feature(1); store(1); repeat(1); exude(1); issue(1); be made for(1); deliver(1); sound like(1); employ(1); be productive of(1); be filled with(1)

Neighborhood bars [IN] be in(10); be located in(10); be found in(4); be situated in(3); serve(2); protect(2); be populated by(2); be used by(2); reside in(2); exist in(2); cater to(2); operate in(1); belong to(1); exist within(1); be located near(1); be known by(1); appeal to(1); involve(1); be nearby(1); be alongside(1); be housed in(1); service(1); be popular in(1); operate(1); control(1); function(1); be built in(1); be located in(1); be in nearby(1); be frequented by(1); work with(1);

be near(1); be inside(1); establish in(1); encircle(1); reside near(1); belong in(1); found in(1)

Night flight [IN] occur at(14); happen at(11); fly at(4); go at(3); be scheduled for(3); leave at(2); be taken at(2); come at(2); take off at(2); be scheduled during(2); be scheduled at(2); operate at(1); fly during(1); arrive at(1); depart at(1); be in(1); operate(1); begin at(1); occur as(1); travel through(1); behold(1); go during(1); occur during(1); land at(1); happen during(1); move at(1); be seen at(1); fly in(1); happen by(1); involve(1); go into(1); end at(1); benefit from(1); originate in(1); run during(1); travel at(1); go on at(1); start at(1); traverse(1); disappear into(1); run at(1)

Nose drops [FOR] be used in(4); be placed in(4); go in(4); clear(4); medicate(3); go into(3); help(3); heal(3); come from(3); fall from(2); be applied in(2); moisturize(2); treat(2); be inserted into(2); be put in(2); be used for(2); soothe(2); be inhaled through(1); run into(1); be meant for(1); drip from(1); penetrate(1); unstop(1); be produced by(1); be dropped in(1); act in(1); unclog(1); drop into(1); be beneficial to(1); irritate(1); clear up(1); decongest(1); run from(1); be inserted in(1); spray into(1); comfort(1); be applied through(1); unstuff(1); act within(1); be made for(1); be for(1); drain from(1); be released from(1); relieve(1); take in(1); unplug(1); be intended for(1); be administered in(1); act on(1); administer to(1); serve for(1); affect(1)

Novelty item [BE] be(13); be called(4); be classified as(3); be seen as(2); be considered(2); represent(2); have(2); resemble(2); be sold as(2); inspire(2); appear(2); display(1); encourage(1); find(1); be perceived as(1); generate(1); reek of(1); be categorized as(1); contain(1); become(1); be known for(1); be loved for(1); be regarded as(1); present(1); seem(1); consist of(1); be marketed as(1); require(1)

Nut bread [HAVE₁] contain(21); include(10); be made with(9); have(8); be made from(5); use(3); be made using(3); feature(2); be filled with(2); taste like(2); be made of(2); come from(2); consist of(2); hold(1); be composed of(1); be blended with(1); be created out of(1); encapsulate(1); diffuse(1); be created with(1); be flavored with(1); incorporate(1); be created from(1); be prepared with(1); sink under(1); comprise(1); eat up(1); be made out of(1); wreck(1); be baked using(1); cover over(1); improve with(1); taste of(1); be baked with(1); rise above(1); surround(1); be about(1)

Office friendships [IN] start in(6); begin at(5); form in(4); come from(4); be made in(3); occur in(3); begin in(3); be in(3); be formed in(3); occur at(3); flourish in(2); originate in(2); happen at(2); remain in(1); revolve around(1); be created in(1); be enjoyed in(1); be created at(1); stem from(1); flourish at(1); evolve from(1); be bonded by(1); reside in(1); be found in(1); be cultivated in(1); involve(1); be shaped at(1); be grow in(1); result in(1); be kindled in(1); pertain to(1); originate at(1); start(1); form at(1); flourish from(1); blossom in(1); blossom at(1); be developed in(1); start at(1); happen in(1); be spawned in(1); start from(1); thrive at(1)

Oil crisis [ABOUT] involve(9); be caused by(8); be related to(3); concern(2); be about(2); revolve around(2); be precipitated by(2); relate to(2); depend on(2); need(1); result from(1); demand(1); center around(1); depend upon(1); occur because of(1); cry for(1); stem from(1); be due to(1); burroughs around(1); terminate in(1); be created from(1); concerns(1); overuse(1); squeeze(1); originate with(1); emerge from(1); be created by(1); derive from(1); result due to(1); refer to(1); deal with(1); contain(1); pertain to(1); be to do with(1); happen due to(1); be triggered by(1); feature(1); scarcens(1); implicate(1); lose(1); be from(1); be effected by(1); affect(1)

Oil well [FOR] contain(15); produce(10); pump(5); provide(4); have(3); draw(3); hold(2); gush(2); spew(2); release(2); extract(2); bring up(2); dispense(2); generate(2); include(1); find(1); supply(1); target(1); be drilled for(1); store(1); be filled with(1); seek(1); pump up(1); mine(1); har-

ness(1); hope for(1); give forth(1); regulate(1); give out(1); be pumped for(1); explode with(1); be on top of(1); access(1)

Olive oil [FROM] come from(19); be made from(11); be pressed from(6); be derived from(5); contain(3); originate from(3); be made of(3); taste like(2); use(1); be(1); be squeezed from(1); be produced from(1); smell of(1); be extracted from(1); run from(1); be composed of(1); smell like(1); be found in(1); process out of(1); be pressed(1); be processed from(1); be pressed out of(1); be made out of(1); emerge from(1); be comprised of(1); be squeezed out of(1); be generated from(1); flow out of(1); be crushed from(1); flow from(1); be made with(1); be expressed from(1)

Onion tears [CAUSE₂] be caused by(18); come from(7); result from(5); be created by(3); react to(2); derive from(2); be induced by(2); be generated by(2); come with(1); be increased by(1); be derived from(1); be sourced by(1); be enduced by(1); be instigated by(1); drip over(1); exude from(1); be due to(1); resemble(1); fall from(1); be produced by(1); begin with(1); be encouraged by(1); be attributed to(1); emerge from(1); be made by(1); be ascribed to(1); develop around(1); be stimulated by(1); be provoked by(1); flow with(1); be associated with(1); be started by(1); be exacerbated by(1); be triggered by(1); commence around(1); arise due to(1); drain under(1); be shed due to(1); be worsened by(1); fall because of(1)

Paper money [MAKE₂] be made of(16); be made from(13); be printed on(6); be(5); be manufactured from(5); consist of(4); be composed of(4); resemble(3); be comprised of(3); feel like(2); fold like(2); be created from(2); look like(2); contain(2); come from(2); be represented by(1); originate from(1); be created with(1); be shaped like(1); be produced from(1); be made out of(1); be created by(1); be constituted by(1); be formed from(1); be produced on(1); signify(1); appear(1); be produced using(1); be formed of(1)

Parent organization [BE] be made up of(6); include(5); be composed of(4); involve(4); consist of(4); contain(4); be for(3); be run by(2); act like(2); concern(2); oversee like(2); affect(2); encourage(1); be supported by(1); have(1); focus on(1); be(1); benefit(1); be created by(1); deal with(1); be funded by(1); serve as(1); supervise like(1); revolve around(1); invite(1); delegate(1); resemble(1); be organized by(1); organize for(1); discuss(1); substitute for(1); be filled by(1); be made for(1); be characterized as(1); be open to(1); be denoted as(1); be marked as(1); be membered by(1); begin as(1); be marketed to(1); be filled with(1)

Party members [HAVE₂] belong to(13); attend(5); join(4); go to(4); represent(2); form(2); comprise(2); be in(2); work for(2); participate in(2); be at(2); vote for(2); make up(2); be involved with(2); be of(1); come with(1); constitute(1); be associated with(1); come from(1); support(1); be involved in(1); campaign for(1); visit(1); be aligned with(1); be invited to(1); think like(1); belong with(1); be included in(1); be affiliated(1); be found in(1); supply(1); affiliate with(1); follow(1); commit to(1); participate with(1); donate to(1); be found at(1); interact at(1); subscribe to(1); belong in(1); descend from(1)

Peanut butter [FROM] be made from(18); come from(9); contain(7); taste like(6); be made of(5); be produced from(3); consist of(3); be derived from(2); include(2); use(2); be composed of(2); be created from(2); be processed from(2); be pressed from(2); be crushed from(2); be from(2); smell like(1); resemble(1); be made with(1); originate from(1); be manufactured from(1); be extrapolated from(1); be formulated from(1); be devised from(1); be made out of(1); be comprised of(1); derive from(1); have inside(1); be developed from(1); be based on(1); envelop(1); be ground from(1); be used with(1); get from(1); be smooshed with(1); require(1)

Pedal extremities [BE] be(6); include(3); use(3); push(3); resemble(2); move(2); be used to(2); look like(2); contain(2); be situated on(2); sport(1); be constructed of(1); maneuver(1); sprout(1); si-

mulate(1); rotate(1); appear like(1); found at(1); carry(1); crank(1); terminate in(1); rest on(1); be placed on(1); act as(1); be found on(1); occur on(1); describe(1); be comprised of(1); operate(1); belong on(1); control(1); refer to(1); relate to(1); be located on(1); be feet and(1); move like(1); work with(1); emulate(1); power(1); manipulate(1); be from(1); require(1)

People power [HAVE₂] come from(12); belong to(5); be used by(3); be generated by(3); be derived from(2); involve(2); be produced by(2); be given to(2); be possessed by(2); reside with(1); encourage(1); include(1); engage(1); emanate from(1); be exercised by(1); reside in(1); drive(1); exist for(1); arise from(1); motivate(1); move(1); emerge from(1); be for(1); be made by(1); originate in(1); be within(1); unite(1); result from(1); flow from(1); be owned by(1); empower(1); originate from(1); be due to(1); be shared by(1); found within(1); occur due to(1); encompass(1); live through(1); refer to(1); be taken by(1); emerge through(1); energize(1); be shown by(1); multiply with(1); grow from(1)

Pet families [HAVE₁] have(14); own(13); love(8); care for(6); consist of(5); like(4); enjoy(4); be composed of(4); contain(4); include(2); keep(2); look after(2); be made of(2); possess(2); nurture(2); be made up of(2); live with(1); adopt(1); house(1); be(1); watch(1); welcome(1); host(1); comprise of(1); be formed from(1); collect(1); believe in(1); take care of(1); be disposed toward(1)

Pet spray [FOR] be used on(9); be for(6); be used for(6); be applied to(5); be made for(3); protect(2); medicate(2); be sprayed on(2); repel(2); be applied on(2); go on(2); cover(2); be designed for(2); deter(1); run from(1); be on(1); be made from(1); be composed of(1); be meant for(1); go onto(1); be put on(1); be distributed by(1); use(1); buy for(1); insects from(1); pass through(1); be exuded by(1); safeguard(1); enter(1); emerge from(1); treat(1); be dispersed on(1); relate to(1); shelter(1); territorialize(1); pertain to(1); come from(1); help(1); be offensive to(1); be intended for(1); consist of(1); help out(1); touch(1); be engaged in by(1); aid(1); deodorize(1); clean(1); deoderizes(1)

Pet theory [BE] involve(6); concern(6); be about(6); pertain to(3); be(2); apply to(2); refer to(2); relate to(2); be like(2); revolve around(1); relate to(1); include(1); center around(1); be cared for like(1); concentrate on(1); resemble(1); be favored like(1); take after(1); be treated like(1); emerge from(1); study(1); be preferred like(1); come from(1); explain about(1); be relevant to(1); regard(1); explain(1); be considered(1); be regarding(1); talk about(1)

Phantom limb [BE] be(10); feel like(10); be like(4); seem(3); act like(3); look like(2); resemble(2); appear(2); exist as(2); have(1); be sensed like(1); behave like(1); mimic(1); feel(1); be called(1); remind of(1); seem like(1); perform like(1); appear like(1); be labelled(1); emulate(1)

Picture album [FOR] contain(22); hold(16); display(11); have(5); protect(4); show(3); store(3); organize(3); be filled with(2); show off(2); be comprised of(1); include(1); keep(1); use(1); be composed of(1); be for(1); be used for(1); develop out of(1); conserve(1); showcases(1); envelop(1); enclose(1)

Picture book [HAVE₁] contain(19); have(10); include(6); use(4); be composed of(3); be made up of(3); display(3); consist of(3); utilize(3); hold(2); show(1); need(1); be(1); be illustrated with(1); incorporate(1); rely on(1); employ(1); present(1); feature(1); portray via(1); contains(1); promote(1); be made of(1); possess(1); accommodate(1); give space to(1); have within(1); be comprised of(1); showcases(1); exhibit(1); organize(1); enclose(1); store(1); be made with(1); require(1)

Pine tree [BE] be(6); have(5); smell like(4); be called(4); produce(4); be made of(3); smell of(3); be made from(2); grow(2); be classified as(2); be composed of(1); be categorized as(1); belong to(1); be named(1); hold(1); grow as(1); be known as(1); recall(1); be labelled as(1); be described as(1); be referred to as(1); classify as(1); relate to(1); bear(1); be categorized as(1); contain(1);

pertain to(1); be known as(1); make(1); define(1)

Plant food [FOR] feed(14); nourish(14); be for(8); sustain(5); be given to(5); be fed to(3); be used by(2); benefit(2); encourage(1); supply(1); be made of(1); be grown from(1); maintain(1); be made for(1); stimulate(1); energize(1); fertilize(1); be meant for(1); be given(1); be necessary to(1); invigorate(1); be eaten by(1); rejuvenate(1); enhance(1); nutures(1); replenishes(1); be taken by(1); be absorbed by(1); be supplied by(1); be consumed by(1); create for(1); istaken up by(1); be used on(1); nurishes(1); aid(1)

Plastic toys [MAKE₂] be made of(18); contain(11); be made from(7); be composed of(6); consist of(4); be manufactured from(3); come from(3); be constructed of(2); use(2); be constructed from(2); be(2); be made out of(2); be built of(2); be molded from(2); utilize(2); have(1); be formed from(1); be derived from(1); be formulated of(1); be created out of(1); be created with(1); be cast in(1); involve(1); be stamped out of(1); be produced from(1); be made using(1); be manufactured with(1); be comprised of(1); be in(1); be made up of(1); be designed out of(1)

Plum wine [FROM] be made from(19); contain(7); taste like(7); come from(7); taste of(2); include(2); be manufactured from(2); be made of(2); need(2); be produced from(2); consist of(2); be made with(2); require(2); be derived from(1); be composed of(1); be distilled from(1); recall(1); be found in(1); be sweet like(1); be formulated from(1); be squeezed from(1); be constituted of(1); emerge from(1); derive from(1); be crafted from(1); be fermented from(1); be generated by(1); be produced using(1); be vinted with(1)

Pole height [HAVE₂] measure(4); describe(3); define(3); be reached by(2); reach(2); refer to(2); come from(2); be set for(1); originate from(1); belong to(1); include(1); dimensionalizes(1); limit(1); be of(1); have(1); terminate at(1); be measured on(1); occur within(1); involve(1); be related to(1); of(1); be approximated by(1); be referenced like(1); be for(1); be made by(1); be exemplified by(1); be ascribed to(1); compare to(1); determine(1); compose(1); be like(1); be compared to(1); be measured for(1); reach to(1); talk about(1); be achieved by(1)

Policy matters [ABOUT] involve(7); concern(7); be related to(4); relate to(4); pertain to(4); deal with(3); refer to(3); be about(3); regard(2); discuss(2); result from(2); reference(2); effect(1); reflect(1); be of(1); have(1); be decided by(1); set(1); define(1); affect(1); require(1); govern(1); be created as(1); be written in(1); use(1); create(1); be developed as(1); clarify(1); follow(1); incorporate(1); be answered by(1); detail(1); comprise(1); enforce(1); derive from(1); be relative to(1); classify(1); belie to(1); be dealt with in(1); be settled by(1); be limited to(1); cover(1); defend(1); go toward(1); be decided as(1)

Pork suet [FROM] come from(11); be made from(10); be made of(8); contain(5); be derived from(3); taste like(3); be rendered from(3); consist of(3); originate from(2); be composed of(2); use(2); be produced from(2); be comprised of(2); include(1); be taken from(1); drain out of(1); be found in(1); be trimmed from(1); substitute for(1); be(1); be related to(1); be mixed from(1); be found on(1); be obtained from(1); exhibit(1); make of(1); con tains(1); be from(1); taste of(1)

Pot high [CAUSE₂] come from(12); be caused by(8); result from(6); be derived from(4); emerge from(3); be created by(2); be induced by(2); be due to(2); involve(2); require(2); be felt from(1); derive from(1); be started by(1); depend on(1); be found in(1); utilize(1); rely on(1); need(1); be achieved with(1); be from(1); come after(1); be created from(1); be produced from(1)

Pressure cooker [USE] use(16); cook with(12); utilize(10); create(5); involve(4); cook by(3); cook under(3); employ(3); require(3); build up(2); generate(2); cook use(2); maintain(1); produce(1); release(1); encourage(1); include(1); harness(1); work on(1); need(1); work under(1); cook via(1); shake from(1); operate use(1); contain(1); necessitate(1); work with(1); heat under(1)

Price dispute [ABOUT] be about(12); involve(10); concern(8); be over(7); question(3); regard(2); argue(2); center on(2); stem from(2); argue about(2); relate to(2); arise over(1); include(1); increase(1); revolve around(1); disagree on(1); debate(1); be concerning(1); examine(1); argue over(1); arise from(1); refer to(1); lower(1); deal with(1); happen because of(1); contest(1); consider(1); be caused by(1); arise due to(1); result from(1); reference(1); object to(1); haggle over(1)

Queen bee [BE] be(13); act as(3); be called(2); rule as(2); serve as(2); reign as(2); act like(2); be positioned as(1); be relegated(1); develop into(1); be recognized as(1); be born as(1); look like(1); come from(1); rule like(1); breed as(1); be classified as(1); evolve into(1); be seen as(1); grow into(1); behave as(1)

Radio communication [USE] use(7); come from(7); need(3); come by(3); be broadcast by(2); be made by(2); be carried by(2); emanate from(2); be transmitted by(2); pass through(2); happen over(2); utilize(2); be transmitted over(1); involve(1); be done over(1); come over(1); occur over(1); rely on(1); occur with(1); flow through(1); be played over(1); depend upon(1); travel through(1); be channeled through(1); be produced by(1); be issued from(1); be conveyed by(1); be received by(1); be broadcast through(1); transmit to(1); be understood across(1); be accomplished through(1); emerge from(1); be carried through(1); be provided by(1); be made with(1); pass between(1); be transmitted using(1); be sent by(1); be achieved by(1); come through(1); be transmitted through(1); originate from(1); emit from(1); be by(1); be performed with(1); transmit by(1); be heard on(1); be created by(1); be generated by(1); be sent by(1); consist of(1); involve use(1); occur via(1); be emitted by(1); be made via(1); require(1); be played by(1); be delivered through(1)

Rice paper [FROM] be made of(15); be composed of(8); be made from(6); contain(5); be created from(4); consist of(4); be produced from(3); come from(3); be made out of(2); be comprised of(2); be made up of(2); make from(1); originate from(1); be derived from(1); include(1); be produced with(1); in make from(1); emanate from(1); be found in(1); use(1); be constructed from(1); be pressed out of(1); look like(1); be formed from(1); make of(1); be fashioned from(1); be produced using(1); be from(1)

Rye whiskey [FROM] be made from(18); come from(7); contain(5); be made of(5); be brewed from(4); be distilled from(3); use(3); be made with(2); be derived from(2); be formulated from(2); be created from(2); be made using(2); derive from(2); consist of(2); taste of(2); be(1); include(1); be produced with(1); smell like(1); taste like(1); be flavored with(1); be produced from(1); ferment(1); emerge from(1); be manufactured with(1); be comprised of(1); be called(1); be composed of(1); be based on(1); require(1)

Salt lake [HAVE₁] contain(19); have(16); include(4); be made of(3); be composed of(3); taste like(3); concentrate(2); taste of(2); be made from(2); hold(2); possess(2); be filled with(2); produce(1); be saturated with(1); provide(1); evaporate to(1); retain(1); be made up of(1); be heavy with(1); be seeped with(1); feature(1); be high in(1); incorporate(1); be rich in(1)

Sand dune [MAKE₂] be made of(17); consist of(6); contain(5); be composed of(4); be comprised of(4); be made from(3); be created from(3); be formed by(2); be formed of(2); be found in(2); be built of(2); be built from(2); be made up of(2); pile up(2); be from(2); be filled with(2); have(1); be(1); be formed from(1); result from(1); originate from(1); shift in(1); be surrounded by(1); need(1); be found among(1); be built by(1); start as(1); amass(1); be in(1); be formed in(1); drift(1); build from(1); employ(1); come from(1); be composed from(1); descend from(1); require(1)

Sap tree [MAKE₁] produce(17); have(7); exude(5); contain(5); ooze(3); give(3); make(3); supply(2);

create(2); secrete(2); drain out(1); be streaked with(1); eject(1); emit(1); leak(1); drip(1); ooze out(1); excrete(1); generate(1); leach out(1); grow(1); run with(1); bleed out(1); bleed into(1); bleed(1); expel(1); manufacture(1); flow with(1); give out(1); look like(1); release(1); be harvested for(1); create sweet(1); make useable(1); abound in(1)

Satellite nation [BE] act like(5); be(4); behave like(3); consist of(3); have(2); use(2); be like(2); resemble(2); be considered(2); be governed as(1); be called(1); depend on like(1); be composed of(1); supply(1); function as(1); be run as(1); orbit like(1); involve(1); view by(1); possess(1); receive from(1); own(1); act as(1); be regarded(1); emerge from(1); generate from(1); be made up of(1); look like(1); operate like(1); serve as(1); revolve like(1); rely on(1); be classified as(1); know about(1); employ(1); seem(1); emulate(1); exist as(1)

Sea breeze [FROM] come from(22); blow from(4); smell like(4); be near(3); blow off(2); originate from(2); come off of(2); emerge from(2); be caused by(2); rise from(1); sweep over(1); smell of(1); flow from(1); come over(1); come off(1); be derived from(1); flow off of(1); waft from(1); be found near(1); emanate from(1); be located near(1); be felt at(1); exist at(1); pass over(1); develop from(1); involve(1); blow along(1); be produced by(1); drift in from(1); blow at(1); generate from(1); go to(1); move over(1); be off(1); begin over(1); originate on(1); approach(1); cool(1); be from(1); cross(1)

Servant girl [BE] work as(17); be(14); act as(6); be employed as(5); serve as(3); be used as(2); be paid as(2); act like(2); be considered(2); have(1); become(1); be kept as(1); represent(1); know(1); function as(1); be described as(1); labor as(1); exemplify(1); talk to(1); look like(1); be born as(1); come from(1); appear(1); work with(1); live like(1); perform as(1)

Sex scandal [ABOUT] involve(15); be about(9); concern(4); revolve around(3); include(3); come from(3); be caused by(3); relate to(2); deal with(2); contain(2); expose(2); talk about(2); have(1); focusses on(1); center around(1); tear down by(1); be due to(1); be made because of(1); be of(1); use(1); focus on(1); speculate about(1); be because of(1); be created by(1); generate from(1); refer to(1); happen because of(1); embody(1); be rumored about(1); be brought about by(1); employ(1); result from(1); entail(1); exploit(1); be concerned with(1); require(1); sensationalizes(1)

Shock treatment [USE] use(16); utilize(7); involve(6); consist of(6); cause(3); include(2); induce(2); give(2); create(2); be comprised of(2); contain(2); entail(2); require(2); provide(1); need(1); produce(1); effect(1); administer(1); elicit(1); be includes(1); be produced by(1); help by(1); primary use(1); incorporate(1); conduct(1); deploy(1); involves(1); relate to(1); be based on(1); rely on(1); employ(1); work by(1); bring about(1)

Sister node [BE] be(4); have(3); be like(3); act like(2); be described as(2); act as(2); be related like(2); accompany(2); belong to(1); hook up with(1); be of(1); resemble(1); supply(1); appear like(1); be related to(1); behave as(1); duplicate(1); be joined by(1); refer to(1); be called(1); pertain to(1); work with(1); be classified as(1); be near(1); make(1); nourish like(1); reside beside(1); be likened to(1); replicate(1)

Smoke signals [USE] be made of(11); use(8); be made from(5); consist of(5); originate from(2); be composed of(2); resemble(2); emerge from(2); be comprised of(2); be made up of(2); look like(2); come from(2); result from(2); be(1); contain(1); rise from(1); be give with(1); be sent with(1); be formed by(1); utilize(1); utilise(1); involve(1); be created from(1); be visible as(1); be encoded as(1); arise from(1); be made out of(1); be made by(1); be read in(1); be generated from(1); result in(1); rise within(1); be generated by(1); advertise through(1); be caused by(1); vanish like(1); utilize(1); be seen(1); make(1); be emitted by(1); be made with(1); signal(1); communicate through(1)

Snow blindness [CAUSE₂] be caused by(24); come from(7); result from(7); be created by(3); be derived from(2); emerge from(2); be induced by(2); be related to(2); come about from(1); be spawned by(1); exemplify(1); descends from(1); be instigated by(1); happen with(1); bring on by(1); happen during(1); bring about by(1); be produced by(1); form from(1); relate to snow(1); happen in(1); be effected by(1); come on in(1)

Sob story [CAUSE₁] cause(12); make(8); create(7); induce(6); lead to(6); produce(5); encourage(4); elicit(4); evoke(4); generate(3); cause to(3); involve(2); result in(2); inspire(2); inflict(1); invite(1); appeal to(1); make for(1); be made of(1); expect(1); recruit(1); fish for(1); draw(1); provoke(1); illicit(1); bring to(1); stimulate(1); embody(1); deserve(1); inswrigates(1); engender(1)

Soldier ant [BE] act like(12); behave like(6); work like(6); be(5); act as(4); resemble(4); imitate(3); emulate(3); be called(2); fight like(2); work as(2); mimic(2); be named for(1); march on like(1); be disciplined like(1); play(1); seem like(1); protect as(1); march for(1); copy(1); behave as(1); march like(1); take after(1); like(1); gobble on(1); defend as(1); serve as(1); appear(1); be like(1); be classified as(1); perform duties of(1); perform as(1)

Song bird [MAKE₁] sing(22); chirp(4); warble(4); make(4); create(3); produce(3); be known for(2); generate(2); have(1); perform(1); provide(1); be identified by(1); burst into(1); utter(1); be famous for(1); tweet like(1); vocalize(1); communicate in(1); erupt in(1); deliver(1); chant(1); croon(1); chortle(1); sound like(1); be praised for(1); tweet(1); love(1); communicate with(1)

Sports activities [BE] involve(12); include(7); be(4); be related to(4); use(3); revolve around(2); be comprised of(2); contain(2); play(1); focus on(1); be associated with(1); be based on(1); support(1); consist of(1); require(1); originate from(1); be composed of(1); center around(1); work on(1); be found in(1); begin with(1); go into(1); have within(1); emerge from(1); partake in(1); be categorized as(1); come from(1); depend on(1); be located in(1); utilize(1); rely on(1); play with(1); be concerned with(1); concern(1); be about(1)

Sports magazine [ABOUT] discuss(9); cover(6); feature(5); be about(5); write about(4); talk about(4); tell about(3); report on(2); involve(2); look at(2); focus on(2); deal with(2); review(2); revolve around(2); detail(2); highlight(2); photograph(2); come from(2); be concerned with(2); concern(2); include(1); promote(1); write on(1); investigate(1); be written about(1); rely upon(1); be devoted to(1); center around(1); trade on(1); be found in(1); go into(1); describe(1); emerge from(1); refer to(1); contain(1); belabor(1); analyze(1); boast about(1); rely on(1); report about(1); explain(1); be filled with(1); be interested in(1)

Star shape [BE] resemble(18); look like(16); be like(3); appear like(3); emulate(2); describe(1); depict(1); represent(1); form(1); be made from(1); mimic(1); derive from(1); appear behind(1); be based on(1); imitate(1); recall(1); come from(1); go around(1)

Starvation diet [USE] cause(9); use(6); lead to(5); involve(4); result in(4); require(4); come from(3); mimic(2); induce(2); go into(2); be like(2); consist of(2); feature(2); promote(1); be(1); rely upon(1); be next to(1); help through(1); encourage(1); include(1); feel like(1); imitate(1); simulate(1); resemble(1); bring(1); be akin to(1); precedes(1); be related to(1); threaten(1); help in(1); emerge from(1); aid in(1); generate from(1); necessitate(1); amount to(1); allow(1); verge on(1); depend on(1); rely on(1); utilize(1); work with(1); employ(1); emulate(1); work by(1); approach(1); bring about(1)

Steam iron [USE] use(12); produce(7); emit(6); create(5); make(4); utilize(4); work by(3); function with(2); need(2); give off(2); run on(2); work with(2); require(2); spray(1); discharge(1); generate(1); incorporate(1); evolve(1); remove with(1); exude(1); operate with(1); steam(1); be

powered by(1); exhale(1); flatten use(1); spurt(1); press with(1); puff(1); press use(1); release(1); apply(1); employ(1); permeate with(1); let off(1); vent(1); be filled with(1)

Steel helmet [MAKE₂] be made of(17); be made from(7); contain(6); be composed of(5); be manufactured from(4); be made out of(3); be comprised of(3); be fashioned from(3); be constructed from(2); consist of(2); be of(1); be(1); be derived from(1); feel like(1); include(1); be produced with(1); be manufactured of(1); be constructed of(1); protect with(1); be constructed out of(1); integrate(1); be fashioned of(1); have(1); be created of(1); conatins(1); use(1); be created from(1); comprise(1); be made using(1); emerge from(1); be formed from(1); come from(1); be formed with(1); utilize(1); be formed of(1); be made with(1); be molded out of(1); be fabricated from(1); require(1)

Stone tools [MAKE₂] be made of(15); be made from(7); work on(4); be used on(3); be created from(3); be manufactured from(2); cut(2); shape(2); be formed from(2); consist of(2); be carved from(2); be shaped from(1); be fashioned from(1); work with(1); be composed of(1); be derived from(1); be cut from(1); carve away(1); consist(1); create out of(1); be constructed out of(1); smoothen(1); break down(1); chisel(1); be constructed from(1); alter(1); be chisled of(1); compose of(1); trim(1); polish(1); be made out of(1); be comprised of(1); be chisled out of(1); be crafted from(1); be developed from(1); come from(1); shine(1); utilize(1); carve(1); be chiseled from(1); transform(1); be fabricated from(1)

Store clothes [FROM] come from(11); be sold in(10); be bought in(8); be found in(5); be purchased at(5); be purchased in(4); be worn in(4); be bought from(3); be purchased from(3); be sold by(2); be from(2); be bought at(2); be made for(2); be located in(2); be used in(1); be found at(1); be stocked in(1); make(1); be marketed in(1); be featured in(1); be obtained from(1); be displayed in(1); be sold(1); be provided by(1); be bought(1); buy(1)

Student committee [MAKE₂] be made up of(10); be composed of(8); be comprised of(7); involve(6); contain(6); consist of(6); be made of(5); include(4); be run by(4); empower(3); represent(2); be formed from(2); be formed of(2); be formed by(1); benefit(1); select(1); comprise(1); enable(1); be led by(1); support(1); judge(1); affect(1); be concerned with(1); be governed by(1); be composed by(1); be conducted by(1); be steered with(1); use(1); be populated from(1); emerge from(1); be controlled by(1); relate to(1); pertain to(1); be operated by(1); be headed by(1); act for(1); entail(1)

Student friends [BE] be(18); come from(2); help(2); be had by(2); be made by(2); include(1); involve(1); act like(1); live as(1); work as(1); tutor(1); become(1); meet as(1); support(1); be enrolled as(1); be considered(1); belong to(1); attend as(1); pal with(1); be met as(1); learn as(1); tand up for(1); befriend(1); comrise(1); walk with(1); associate with(1); be in(1); enter as(1); generate from(1); study like(1); defend(1); be started as(1); be classified as(1); exist as(1); matriculate as(1)

Student power [HAVE₂] come from(10); belong to(5); be held by(4); be used by(4); be possessed by(4); involve(3); be created by(3); be wielded by(2); be demonstrated by(2); be exercised by(2); emerge from(2); originate with(2); derive from(2); be about(2); lie within(1); enable(1); teach(1); inspire(1); be given to(1); be sought by(1); fill(1); be derived from(1); be given(1); be expressed by(1); devolve from(1); emanate from(1); be of(1); be derived by(1); found within(1); rule(1); be entrusted to(1); emanates from(1); radiate from(1); feed from(1); be attributed to(1); be yielded by(1); be in(1); control(1); emanate(1); relate to(1); be seized by(1); be exerted by(1); be weilded by(1); be owned by(1)

Student problems [HAVE₂] involve(6); concern(5); happen to(5); be experienced by(4); affect(4);

belong to(4); be had by(4); be faced by(3); occur with(2); be caused by(2); bother(2); plague(2); come from(2); worry(2); be related to(2); hurt(1); afflict(1); hinder(1); be made by(1); frustrate(1); emerge in(1); impede(1); cause(1); originate from(1); trouble(1); arise from(1); be encountered by(1); be created by(1); be in(1); harm(1); be solved by(1); result from(1); occur to(1); develop with(1); drain(1); face challenge(1); absorb(1); be had about(1); relate to(1); oppress(1); be associated with(1); burden(1); be owned by(1); log down(1); occur(1); obstruct(1); be suffered by(1); be found in(1); aggravate(1); torment(1); frighten(1); overwhelm(1); encumber(1); be possessed by(1); stress(1); be about(1)

Sugar cube [MAKE₂] be made of(20); contain(9); consist of(9); be made from(7); be composed of(5); be(4); be comprised of(3); be formed from(3); taste like(2); look like(1); be derived from(1); be all(1); be made up of(1); build of(1); dissolve like(1); be constructed of(1); portion(1); be of(1); be formulated from(1); store(1); be from(1); be formed of(1); be pressed(1); encompass(1)

Summer months [IN] occur during(13); be in(9); occur in(8); happen in(7); make up(3); be found in(3); comprise(3); include(2); be during(2); represent(2); contain(2); begin in(1); span(1); appear in(1); come in(1); constitute(1); be considered(1); occur(1); be characterized by(1); pass in(1); be included in(1); come during(1); recall(1); happen during(1); fall in(1); encompass(1); be warm like(1); arrive in(1); happen during(1); consist of(1); be defined as(1); define(1); be over(1); predict(1)

Summer travels [IN] occur during(12); happen in(10); occur in(8); begin in(3); involve(3); be in(3); be taken in(3); occur(2); be done in(2); happen during(2); be during(2); be made in(2); include(1); be taken during(1); begin during(1); be organized for(1); be had during(1); rev up(1); extend over(1); be found in(1); explore(1); go into(1); brighten our(1); enhance(1); be booked for(1); undertake in(1); end in(1); enter into(1); come from(1); be enjoyed during(1); enjoy(1); happen about(1)

Surface tension [IN] be on(5); occur on(5); occur at(4); happen at(3); exist on(2); affect(2); involve(2); reside on(2); happen on(2); be found on(2); shape(2); cover(2); define(2); remain on(1); disrupt(1); form(1); appear on(1); effect(1); impact(1); manifest at(1); retain(1); influence(1); preserve(1); be measured at(1); be due to(1); be related to(1); build on(1); relate to(1); stay on(1); suspend on(1); be derived from(1); harden(1); lie at(1); extant in(1); exude from(1); be of(1); abruptly from(1); fester on(1); lie across(1); involves(1); be experienced on(1); be formed on(1); be created on(1); control(1); be specific to(1); be located on(1); break into(1); come from(1); be felt on(1); form at(1); strengthen(1); tense(1); hold across(1)

Suspense film [CAUSE₁] contain(9); have(7); create(6); cause(6); include(3); generate(3); involve(3); be(2); utilize(2); produce(2); build(1); center around(1); build up(1); use(1); provide(1); carry(1); offer(1); be filled with(1); be cloaked in(1); impart(1); encourage(1); be characterized by(1); purpose be(1); promote(1); exude(1); supply(1); induce(1); need(1); acclaim from(1); descend in(1); emerge from(1); be comprised of(1); be steeped in(1); be marked by(1); result in(1); give(1); deal with(1); inspire(1); showcases(1); foster(1); feature(1); be written for(1); engender(1); project(1); bring about(1); be about(1)

Tape measure [BE] be made of(8); use(5); look like(5); be(3); be on(3); resemble(3); be made from(2); be in(2); consist of(2); have(1); be formed like(1); be composed of(1); be printed on(1); be created out of(1); be formed by(1); flex like(1); be of(1); measure with(1); be manufactured on(1); roll out like(1); form of(1); be determined with(1); be shaped like(1); roll like(1); be made by(1); be flexible like(1); contain(1); wind like(1); run along(1); be like(1); utilize(1); be used with(1); employ(1); be provided by(1); be determined by(1); be written on(1); be marked on(1);

be composed from(1); be quantified using(1); bend like(1); be rolled like(1)

Target structure [BE] be(10); have(4); include(3); contain(3); involve(3); represent(2); resemble(2); provide(2); become(2); be used as(1); hold(1); hit(1); aim at(1); meet(1); house(1); comprise(1); identify(1); reach(1); give(1); look like(1); specify(1); surround(1); be on(1); be made from(1); uphold(1); be aimed at(1); indicate(1); be made(1); be found in(1); supply(1); be made of(1); levy(1); be shot at like(1); be marked as(1); be classified as(1); pinpoint(1); make up(1); aim for(1)

Tax law [ABOUT] be about(9); involve(7); govern(4); describe(4); regulate(3); concern(3); be related to(3); pertain to(3); deal with(2); relate to(2); explain(2); impose(1); change(1); mandate(1); encompass(1); enforce(1); enable(1); address(1); specialize in(1); justify(1); determine(1); collect(1); define(1); require(1); be written about(1); revolve around(1); center on(1); prescribe(1); dictate(1); focus on(1); legalize(1); rfeers to(1); interpret(1); refer to(1); contain(1); touch on(1); concernsn(1); itemize(1); set forth(1); quantify(1); be regarding(1); be concerned with(1)

Teaching profession [BE] involve(12); include(7); require(7); be about(4); promote(2); focus on(2); revolve around(2); center on(2); use(2); espouse(2); utilize(2); concern(2); demand(1); perform(1); do(1); be in(1); study(1); specialize in(1); enjoy(1); encourage(1); be based around(1); be devoted to(1); profess(1); reward(1); engage in(1); embrace(1); be comprised of(1); refer to(1); be based on(1); cherish(1); focus upon(1); foster(1); help with(1); practice(1); consist of(1); be centered on(1); stress(1); go along with(1)

Tear gas [CAUSE₁] cause(18); create(9); make(7); generate(5); produce(3); induce(3); provoke(2); initiate(1); lead to(1); encourage(1); accelerate(1); compel(1); intensify(1); result in(1); stimulate(1); irritate(1); shades(1); incite(1); force(1); engender(1); bring about(1)

Testtube baby [FROM] be conceived in(13); come from(10); be created in(8); be made in(7); begin in(4); be started in(4); originate in(4); be fertilized in(4); originate from(2); start in(2); be formed in(2); gestate in(1); be born from(1); be grown in(1); emanate from(1); be started out in(1); be born in(1); be created with(1); be cultivated in(1); exist in(1); be created from(1); be made using(1); be for(1); first grow in(1); be concieved in(1); live in(1); be designed in(1); be beholden to(1); be formed within(1); be begun in(1); in conceive in(1)

Tire rim [HAVE₂] surround(7); be on(5); hold(4); support(4); fit(2); be in(2); border(2); be of(2); be made for(2); cover(2); circle(2); encircle(2); protect(1); include(1); involve(1); accept(1); connect to(1); belong to(1); be surrounded by(1); be included with(1); balance(1); be required by(1); give support to(1); go around(1); be found in(1); go with(1); outline(1); be related to(1); encompass(1); be found on(1); lock in(1); run through(1); shape(1); be within(1); seal(1); be around(1); be located in(1); be inside of(1); be inside(1); enclose(1)

Tobacco ash [FROM] come from(21); derive from(5); be produced from(4); contain(4); be made of(3); be from(3); be made from(2); be derived from(2); be(2); emerge from(2); result from(2); be composed of(1); be created out of(1); burn from(1); be due to(1); be produced by(1); remain of(1); be created from(1); stick to(1); be created by(1); be dropped from(1); be sloughed from(1); look like(1); be fueled by(1); be generated by(1); be caused by(1); form from(1); be emitted by(1); be shed from(1); descend from(1)

Vacuum cleaner [USE] use(12); clean by(6); create(3); utilize(3); have(2); need(2); suck into(2); work with(2); apply(2); require(2); include(1); be(1); do(1); act as(1); make(1); utillizes(1); sweep up use(1); work on(1); clean with(1); function as(1); involve(1); utilices(1); clean use(1); suck with(1); work via(1); suck in(1); function by(1); order to(1); suck by(1); be categorized as(1); come from(1); pick up with(1); work through(1); employ(1); know as(1); work by(1); work use(1)

Vapor lock [CAUSE₂] be caused by(11); result from(5); contain(4); involve(2); keep out(2); control(2); come from(2); occur because of(1); keep inside(1); inhibit(1); be instigated by(1); be due to(1); develop in(1); build up(1); be found in(1); close of(1); restrain(1); be made of(1); form(1); be produced by(1); repress(1); be(1); be related to(1); regulate(1); hinder(1); emerge from(1); concentrate(1); originate with(1); be created by(1); eliminate(1); hold in(1); arrest(1); seal in(1); develop because of(1); move with(1); happen due to(1); keep in(1); harbor(1); impede(1); block(1); close with(1)

Vegetable soup [HAVE₁] contain(14); be made from(11); have(9); be made of(8); include(5); taste like(4); use(3); consist of(3); be composed of(3); come from(3); be made with(2); call for(1); hold(1); contain be(1); house(1); stew(1); incorporate(1); be created from(1); be made using(1); be supplied with(1); have within(1); be comprised of(1); cook up with(1); be made up of(1); be found with(1); consist of have(1); be cooked with(1)

Voice vote [USE] be made by(9); use(6); be given by(3); be cast using(3); come from(3); be conducted by(2); originate from(2); count(2); come by(2); be counted by(2); need(2); be done by(2); be made using(2); be cast by(2); require(2); have(1); be taken by(1); happen through(1); arrive via(1); occur through(1); listen to(1); be expressed by(1); be shown with(1); emanate from(1); be held by(1); be signified by(1); register with(1); operate on(1); be indicated with(1); be represented by(1); find(1); happen by(1); be registered by(1); be gathered using(1); comprise(1); determine by(1); emerge from(1); be cast with(1); depend on(1); utilize(1); rely on(1); hold by(1)

Wastebasket category [BE] belong in(6); include(4); go in(3); describe(2); come from(2); be about(2); fill(1); belong to(1); be named(1); be put into(1); be filed in(1); simulate(1); resemble(1); have(1); be named after(1); use(1); provide(1); involve(1); focus on(1); be related to(1); act like(1); go into(1); act as(1); be treated like(1); be for(1); classify as(1); deal with(1); contain(1); relegate to(1); serve as(1); utilize(1); be like(1); allude to(1); define(1); emulate(1); trash into(1); explain(1); concern(1)

Water drop [MAKE₂] be made of(16); contain(9); consist of(7); be composed of(6); be(6); be comprised of(3); be formed by(2); come from(2); have(1); emerge from(1); swim in(1); be created by(1); hold(1); be formed from(1); drip(1); stretch(1); work in(1); be formulated from(1); form from(1); come out of(1); be filled with(1)

Water mark [CAUSE₂] be made by(5); be created by(5); indicate(4); be made from(4); come from(4); be caused by(4); look like(3); be left by(3); be made with(3); show(1); intaglio(1); be located near(1); reveal(1); be indicated by(1); be impressed by(1); use(1); be made of(1); appear like(1); involve(1); leave(1); stain like(1); imprint(1); be produced from(1); emboss(1); delineate(1); emerge from(1); stain(1); deform like(1); be generated by(1); appear(1); appear from(1); circumscribe(1); measure(1); level to(1); demarcate(1)

Water wheel [USE] use(8); move(6); be powered by(5); utilize(3); turn by(2); spin(2); work in(2); run on(2); rotate in(2); require(2); turn use(2); turn with(1); be made of(1); carry(1); turn in(1); be operated by(1); power(1); be spun by(1); flow through(1); be composed of(1); utilizes(1); rotate use(1); spin with(1); harness(1); turn from(1); transport(1); spin through(1); reside in(1); be created with(1); transmit(1); be pushed by(1); need(1); activate with(1); exist in(1); be propelled by(1); rotate via(1); tumble(1); turn under(1); flow with(1); necessitate(1); relocate(1); work with(1); apply(1); direct(1); grind with(1); take out(1); roll(1)

Weekend boredom [IN] occur on(15); happen during(9); happen on(9); occur during(5); come on(3); plague(1); happen along(1); coincide with(1); happen through(1); occur through(1); emerge at(1); set in on(1); begin during(1); occur because of(1); result during(1); come by(1); be due

to(1); begin within(1); mar(1); occur within(1); occur over(1); detract from(1); involve(1); be produced by(1); be localized to(1); begin along(1); permeate(1); be experienced on(1); occur along(1); happen within(1); happen near(1); accrue during(1); happen because of(1); spoil(1); lengthen(1); happen at(1); come from(1); be felt on(1); characterize(1); experience in(1); occur at(1); be caused by(1); begin through(1); result from(1); transpire during(1); affect(1); hit during(1)

Winter season [BE] be(7); include(6); occur in(5); be in(5); occur during(4); feel like(2); happen during(2); be called(2); come in(2); exist in(2); encompass(2); describe(2); happen at(2); consist of(2); define(2); happen in(2); coincide with(1); begin in(1); end in(1); cover(1); hold(1); encapsulate(1); represent(1); recall(1); reside within(1); begin with(1); behave as(1); parallel with(1); comprise(1); be comprised of(1); exemplify(1); contain(1); be known as(1); occur at(1); be like(1); emulate(1); mark(1); be classified(1); beckon(1); concern(1)

Winter sports [IN] be played in(17); occur in(9); happen in(7); occur during(5); be played during(4); happen during(3); require(3); be done in(2); celebrate(2); be in(2); rely upon(1); be enjoyed in(1); be conducted in(1); start in(1); be found in(1); be related to(1); be held in(1); be found during(1); be practised in(1); originate in(1); come in(1); correlate with(1); come from(1); signify(1); utilize(1); rely on(1); be in season of(1); be designated for(1); suit(1); exploit(1); be participated in during(1)

Women professors [BE] be(22); teach(2); look like(2); be born(2); research(1); study(1); be comprised of(1); behave like(1); include(1); be gendered(1); dress like(1); identify as(1); come from(1); resemble(1); seem like(1); be classified as(1); emulate(1); exist as(1); educate as(1); live as(1); work as(1)

Wood shavings [FROM] come from(17); be made of(11); be made from(6); be composed of(4); contain(3); be produced from(3); be comprised of(3); be derived from(2); be created by(2); originate from(1); be sheared from(1); be formed by(1); fall from(1); be products of(1); be created of(1); supply(1); be produced by(1); be carved off(1); be created from(1); be peeled from(1); come out of(1); be made out of(1); emerge from(1); be formed from(1); be formed with(1); consist of(1); be chipped from(1); result from(1); remain from(1); be carved from(1)

Worker teams [MAKE₂] be made up of(11); include(10); be composed of(6); be comprised of(6); have(5); consist of(5); contain(4); comprise(3); employ(3); be made from(3); use(2); be made of(2); bring(1); involve(1); incorporate(1); be formed from(1); be staffed by(1); possess(1); be comprised by(1); be derived of(1); unify(1); organize(1); be formed with(1); utilize(1); exploit(1); be organized with(1)

Vecteurs prédictifs pour uniquement la première réponse des formulaires

Abortion problem [ABOUT] involve(6); be related to(3); concern(3); occur with(2); come from(2); occur in(2); be caused by(1); be(1); surround(1); originate from(1); refer to(1); relate to(1); stem from(1); lie with(1); affect(1)

Abortion vote [ABOUT] be about(4); concern(3); support(3); involve(3); decide on(1); prohibit(1); approve(1); be for(1); be against(1); be taken regarding(1); be pro(1); feed(1); define(1); terminate in(1); regulate(1); affect(1)

Accident weather [CAUSE₁] cause(13); promote(2); be conducive to(1); lead to(1); occur by(1); result in(1); precipitate(1); be conducive to(1); foster(1); happen by(1); feed(1); generate(1); create(1); precipitates(1)

Adolescent turmoil [IN] occur in(5); be experienced by(5); affect(4); happen to(3); be caused by(2); involve(2); effect(1); bother(1); occur to(1); emerge from(1); appear in(1); confuse(1); befall(1); traverse(1); afflict(1)

Adventure story [ABOUT] contain(6); describe(3); tell of(3); have(3); be about(3); involve(2); include(1); feature(1); be(1); tell(1); concern(1); be characterized by(1); be found in(1); discuss(1); provide(1)

Air pressure [CAUSE₂] come from(6); be caused by(5); occur in(1); be exerted by(1); use(1); force(1); be measured in(1); effect(1); be made by(1); be induced by(1); relate to(1); be caused by(1); be due to(1); be measured from(1); involve(1); be exhibited by(1); be filled with(1)

Alligator leather [FROM] be made from(11); come from(6); be made of(2); be on(1); belong to(1); be obtained from(1); be(1); be skin from(1); be from(1); be born of(1); be taken from(1)

Apple cake [HAVE₁] be made from(6); be made with(5); contain(4); be made of(3); taste like(2); come from(2); use(1); originate from(1); be composed of(1); be baked with(1); be filled with(1)

Apple core [HAVE₂] come from(10); be in(3); be from(2); be located in(1); be contained within(1); be found in(1); be inside of(1); center(1); be inside(1); be within(1); be of(1); reproduce(1)

Arms budget [FOR] be for(5); pay for(3); include(2); concern(2); be earmarked for(1); be meant for(1); be used for(1); be allocated to(1); relate to(1); consider(1); cover(1); provide(1); fund(1); be set aside for(1); provide for(1); be spent on(1); be utilized for(1); buy(1)

Automobile plant [FOR] manufacture(12); make(10); build(2); produce(1); assemble(1); create(1)

Bacon grease [FROM] come from(17); be derived from(2); be made from(2); be rendered from(1); stick to(1); be exuded by(1); derive from(1); make(1); be produced by(1); be fried out of(1)

Basketball season [FOR] involve(4); feature(4); include(3); contain(3); be for(2); have(2); demonstrate(1); be devoted to(1); welcome(1); focus on(1); be active for(1); encompass(1); affect(1); to play(1); highlight(1)

Bear country [HAVE₁] have(8); contain(6); harbor(3); nourish(2); be inhabited by(2); feed(1); nurture(1); be populated with(1); love(1); be home to(1)

Beehive hairdo [BE] look like(11); resemble(9); be shaped like(3); emulate(1); rise like(1)

Birth pains [CAUSE₂] be caused by(6); occur during(3); come from(2); accompany(2); happen during(2); precede(1); come before(1); occur with(1); coincide with(1); emit from(1); happen with(1); be due to(1); be had during(1); occur at(1); complicate(1); be from(1)

Blanket excuse [BE] cover like(8); act like(3); be like(2); be used as(1); behave like(1); serve to(1);

be analogous to(1); be(1); be given as(1)

Bronze statue [MAKE₂] be made of(14); be cast from(2); be made from(1); be composed of(1); be(1); be made out of(1); be constructed from(1); be manufactured from(1); contain(1); be fabricated from(1); come from(1)

Budget speech [ABOUT] be about(8); discuss(4); concern(3); be on(1); describe(1); turn in(1); refer to(1); be made on(1); detail plan for(1); pertain to(1); explain(1); talk about(1)

Bull ring [FOR] contain(8); house(2); be worn by(2); be in(1); be located in(1); be on(1); be for(1); feature(1); fence in(1); hold(1); have(1)

Cable network [MAKE₂] consist of(3); run on(2); use(2); provide(2); carry(2); be on(1); run through(1); be connected by(1); be shown on(1); be carried on(1); connect(1); have(1); utilize(1); be in(1); be relayed by(1); be made up of(1); program(1); be supplied by(1); reach through(1); be played on(1); broadcast use(1); be formed of(1)

Cactus plant [BE] be(12); look like(2); be composed of(1); derive from(1); relate to(1); bear(1); contain(1); be classified as(1); be categorized with(1); be related to(1); birth(1); be described as(1); grow from(1)

Candy cigarette [MAKE₂] be made of(13); taste like(4); be made from(4); make from(1); consist of(1); be(1); look like(1); be made with(1); resemble(1)

Cane sugar [FROM] come from(14); be made from(4); be processed from(2); be created from(2); be derived from(1); be made of(1); be produced from(1)

Canine companion [BE] be(20); derive from(1); assist as(1); descend from(1); be of(1)

Cash basis [BE] be based on(3); use(3); involve(3); come from(2); include(1); consist of(1); be backed by(1); depend on(1); determine(1); rely on(1); be made of(1); begin with(1); accept(1); be related to(1); be regarding(1); require(1); be concerned with(1)

Ceiling price [BE] hit(5); be(3); reach(2); form(2); go to(2); be at(2); constitute(1); serve as(1); be like(1); be terminated at(1); maxes at(1); at(1)

Child actor [BE] be(22); involve(1); portray(1); be classified as(1)

Childhood dreams [IN] occur in(5); occur during(3); come in(2); originate in(2); come from(1); have(1); start in(1); be dreamt in(1); revolve around(1); be had in(1); happen during(1); invade(1); come about in(1); happen(1); be spawned in(1); happen in(1); be about(1)

Chocolate bar [MAKE₂] be made of(11); be made from(9); contain(6); be composed of(2)

Cigarette burn [CAUSE₂] be caused by(13); come from(6); be made by(2); result from(1); resemble(1); be done with(1); be left by(1)

Cigarette war [ABOUT] involve(10); revolve around(2); be over(2); be about(2); ban(1); be fought because of(1); be fought over(1); be against(1); concern(1); oppose(1)

Citizen soldier [BE] be(19); protect(1); represent(1); be classified as(1); be selected from(1); live as(1); come from(1)

City folk [IN] live in(17); come from(3); reside in(3); emerge from(1)

City wall [HAVE₂] surround(20); protect(2); enclose(2); contain(1); encircle(1); be in(1); border around(1); surrounds(1)

Coal dust [FROM] come from(15); be(2); be generated from(1); be attached to(1); be made from(1);

be composed of(1); be spread by(1); emerge from(1); be made of(1); derive from(1); be produced by(1); be produced from(1)

Coffee nerves [CAUSE₂] be caused by(9); come from(5); be from(2); be caused from(1); run on(1); emanate(1); be affected by(1); be blamed on(1); recognize(1); be generated by(1); need(1); be jangled from(1); be frayed from(1)

Coke machine [FOR] dispense(15); sell(3); vend(2); distribute(2); serve(1); give(1); contain(1); offer(1)

College town [HAVE₁] contain(9); have(8); support(1); host(1); be attached to(1); accomodates(1); begin because of(1); nourish(1); sorround(1); encompass(1)

Collie dog [BE] be(10); be called(2); look like(1); come from(1); resemble(1); belong to(1); emerge from(1); be named(1); be implied to(1); be bred from(1); exhibit(1); harbor(1); be classified as(1)

Color television [HAVE₁] display in(5); be in(4); display(3); show(2); contain(2); project in(2); show in(2); use(1); transmit in(1); include(1); project with(1); broadcast in(1); illuminate(1); utilize(1); employ(1)

Communist tenet [IN] be held by(5); be believed by(3); support(2); be followed by(1); be believed in by(1); belong to(1); be made by(1); be created by(1); be defined by(1); come from(1); be of(1); agree with(1); converse with(1); be upheld by(1); attract(1); purport(1); concern(1)

Company assets [HAVE₂] belong to(10); be owned by(6); be generated by(2); lie within(2); help(1); be(1); be held by(1); come from(1); affect(1); be possessed by(1)

Concrete desert [MAKE₂] be made of(12); be covered in(1); be composed of(1); be comprised of(1); be in(1); include(1); be made up of(1); look like(1); be turned in(1); be fashioned from(1); resemble(1); seem like(1); embody(1); utilize(1); consist of(1); appear like(1)

Concussion force [CAUSE₁] cause(17); create(3); result in(2); produce(1); be found in(1); lead to(1); induce(1); form(1); be produced by(1)

Cooking utensils [FOR] be used for(15); be used in(7); facilitate(1); be made for(1); be found in(1); aid in(1); aid(1); underscore(1)

Copper coins [MAKE₂] be made of(11); be made from(8); be composed of(1); look like(1); contain(1); utilize(1); be made out of(1)

Coriander curry [HAVE₁] contain(8); be made from(5); come from(3); be garnished with(2); have(2); be flavored with(1); produce(1); feature(1); look like(1); be cooked with(1)

Country butter [FROM] be made in(11); come from(8); be produced in(2); be from(1); taste(1); come out of(1); remind of(1)

Country visitors [FROM] come from(6); go to(5); visit(4); be from(3); hail from(1); arrive from(1); visit from(1); come to(1); go in(1); originate in(1); travel to(1); holiday in(1)

Cream sauce [HAVE₁] contain(9); be made of(5); be made from(2); be like(1); use(1); be composed of(1); be shaped as(1); come from(1); be of(1); resemble(1); comtains(1); be thickened with(1); be feels like(1)

Daisy chains [MAKE₂] be made of(11); be made from(6); look like(2); be made up of(1); contain(1); cut(1); resemble(1); link(1); emulate(1); be created from(1); incorporate(1); be made with(1); loop like(1)

Deficiency disease [CAUSE₂] be caused by(12); come from(2); result in(1); stem from(1); be asso-

ciated with(1); have(1); involve(1); result from(1); cause(1); arise from(1); be caused from(1); be created by(1); refer to(1); occur due to(1); affect(1)

Desert rat [IN] live in(21); be located in(1); be found in(1); inhabitates(1); dwell in(1); come from(1)

Disaster flick [ABOUT] be about(4); involve(3); describe(2); show(2); portray(2); be written about(1); be created by(1); center around(1); be based on(1); simulate(1); come from(1); feature(1); be(1); concern(1)

Disease germ [CAUSE₁] cause(18); carry(3); spread(2); produce(1); generate(1)

Drug deaths [CAUSE₂] be caused by(12); be due to(4); result from(3); occur because of(2); happen because of(2); occur from(1); be from(1); emanate from(1)

Enemy strength [HAVE₂] be possessed by(6); belong to(4); be attributed to(1); be controlled by(1); be held by(1); arise out of(1); come from(1); characterize(1); be of(1); evaluate(1); have(1); build up(1); be found in(1); be exerted by(1); be made of(1); define(1); carry(1); encompass(1)

Energy emergency [ABOUT] involve(7); be due to(3); concern(3); relate to(2); be from(2); require(1); impact(1); restrict(1); refer to(1); pertain to(1); be caused by(1); be regarding(1); be about(1)

Evening hours [IN] occur in(9); be in(7); occur during(3); start in(1); include(1); be considered(1); fall during(1); indicate(1); pertain to(1); happen in(1); encompass(1); occur at(1)

Extension ladder [HAVE₁] have(11); utilize(2); provide(2); contain(1); allow(1); form(1); include(1); pull out(1); perform(1); supply(1); use(1); add(1); possess(1); be made with(1); be able to make(1)

Faith cure [USE] rely on(5); be caused by(3); involve(2); be done by(2); be related to(2); come from(1); require(1); rely upon(1); occur through(1); be ascribed to(1); occur because of(1); happen from(1); depend on(1); happen by(1); be rooted in(1); be powered by(1)

Family antiques [HAVE₂] belong to(9); be owned by(6); be kept in(1); come from(1); pass through(1); remain with(1); be linked to(1); be passed down within(1); be kept by(1); survive(1); relate to(1); be handed down in(1)

Family problems [IN] involve(4); occur in(2); plague(2); affect(2); be created by(1); include(1); be within(1); be suffered by(1); arise in(1); occur within(1); be found in(1); crop up in(1); happen in(1); belong to(1); be related to(1)

Farm boy [FROM] live on(14); work on(6); come from(4); work in(3); reside on(1)

Fatigue headache [CAUSE₂] be caused by(13); result from(4); come from(3); be brought on by(1); cause(1); occur do to(1); bring on by(1); come from(1)

Field mouse [IN] live in(19); inhabit(2); be in(1); roam around(1); originate in(1); dwell in(1); come from(1)

Financing dilemma [ABOUT] involve(6); be caused by(3); concern(3); be about(3); come from(2); cause(2); emerge from(1); occur in(1); relate to(1); complicate(1); be found in(1); result from(1); affect(1)

Finger cymbals [USE] be worn on(5); be played by(4); be played with(3); go on(2); use(2); attach to(2); be used by(2); wear on(1); involve(1); be(1); look like(1); work on(1)

Finger lakes [BE] resemble(9); look like(6); be shaped like(4); flow in(2); contain(1)

Fish scales [HAVE₂] be on(6); cover(5); be found on(4); grow on(3); come from(2); fall from(1);

protect(1); weigh(1); shed from(1); be in(1); be worn by(1); be layered on(1)

Flounder fish [BE] be(8); look like(2); be sold as(1); be called(1); come from(1); appear(1); struggle like(1); be classified as(1); be seen in(1); love to(1); be related to(1); do(1); be known to(1); seem to(1)

Flu virus [CAUSE₁] cause(17); be(1); carry(1); involve(1); come from(1)

Frog man [BE] resemble(6); look like(4); swim like(4); be like(2); be(2); love(1); emerge from(1); dive like(1); behave like(1); emulate(1); work like(1); seem like(1)

Fruit tree [HAVE₁] bear(18); produce(7); have(2); bare(1); grow(1); bloom(1)

Future shock [CAUSE₂] involve(4); happen in(3); occur in(2); come from(2); be caused by(2); be relative to(1); remain(1); happen during(1); occur within(1); be linked to(1); be dismayed about(1); be from(1); concern(1); be about(1)

Garter snake [BE] resemble(8); look like(5); be called(2); like(1); be thin like(1); be named for(1); wear(1); be banded like(1); crawl through(1); have(1)

Gas stove [USE] use(15); run on(4); burn(2); emit(2); run by(1); be powered by(1); heat by(1); exumes(1)

Glass eye [MAKE₂] be made of(18); be made from(4); be composed of(2); come from(2); be like(1); be constructed of(1); contain(1); be created from(1)

Government employment [IN] be provided by(4); involve(3); be sponsored by(2); be made by(2); be with(2); serve(1); be by(1); be within(1); be given by(1); happen for(1); be offered by(1); begin with(1); benefit(1); be performed by(1)

Government land [HAVE₂] be owned by(14); belong to(9); come from(1)

Grain alcohol [FROM] be made from(10); come from(4); be produced from(3); be distilled from(2); be made of(2); contain(2); be derived from(1); be fermented from(1); incorporate(1); be extracted from(1)

Growth hormone [CAUSE₁] promote(9); cause(4); stimulate(3); regulate(2); induce(1); help in(1); affect(1); aid(1); spark(1); supplement(1)

Gutter language [FROM] come from(7); resemble(3); belong in(2); originate from(2); be heard in(1); be in(1); relate to(1); represent(1); pertain to(1); stem from(1); remind of(1); be used in(1); be from(1)

Hairpin turn [BE] resemble(9); look like(7); mimic(2); be like(1); use(1); be sharp like(1); turn as(1); be curved like(1); be lik(1); remind of(1)

Hand brake [USE] be operated by(9); be applied by(3); operate by(3); require(3); be used by(2); use(1); be applied by(1); be operated with(1); go in(1); need(1); be pushed with(1); be handled by(1); be worked by(1)

Handlebar mustache [BE] look like(11); resemble(10); be like(1); appear like(1); recall(1)

Head noun [BE] be(6); involve(2); refer to(2); be at(2); be found in(2); occur at(1); form(1); act as(1); describe(1); represent(1); come at(1); challenge(1); touch(1); lead(1)

Headache pills [FOR] cure(11); relieve(4); treat(3); help with(1); help(1); be for(1); reduce(1); deaden(1); feed(1); negate(1); heal(1)

Heart design [BE] resemble(4); be shaped like(4); include(3); look like(2); incorporate(2); contain(2);

involve(2); use(1); warm(1); be composed of(1); be made from(1); center around(1); relate to(1); be shaped in(1); be shaped as(1); be made of(1); encompass(1)

Heat rash [CAUSE₂] be caused by(24); come from(3); emerge from(1); result from(1); be caused by(1); be from(1)

Hermit crab [BE] act like(7); live like(4); resemble(4); be(3); hide like(1); be recluse like(1); mimic(1); emulate(1); be called(1); look like(1); recall(1)

History conference [ABOUT] discuss(4); focus on(4); be about(4); deal with(3); concern(3); involve(2); study(1); be on(1); be related to(1); pertain to(1); talk about(1)

Home remedy [FROM] be made at(9); come from(4); be used in(3); originate at(2); be done at(1); be used at(1); be found at(1); originate from(1); be made in(1); be provided at(1); help at(1)

Honey bee [MAKE₁] make(16); produce(7); create(2); provide(1); manufacture(1); store(1)

Horse doctor [FOR] treat(15); care for(3); tend to(2); heal(2); serve(1); cure(1); fix(1); practise on(1); work on(1)

Hydrogen bomb [USE] use(6); be made from(4); be made of(4); be composed of(2); utilize(2); contain(2); ignite(1); need(1); be fueled by(1); consume(1); come from(1)

Immigrant minority [MAKE₂] be comprised of(6); be made of(4); be composed of(3); consist of(3); be made up of(3); contain(2); be(1); form(1); have(1)

Infant colonies [BE] be composed of(3); contain(3); resemble(2); be like(2); be(2); behave like(1); be comprised of(1); be populated by(1); be made up of(1); be still(1); be made of(1); involve(1); house(1); compose of(1)

Job tension [CAUSE₂] be caused by(8); result from(4); come from(3); be created by(2); occur on(1); occur in(1); occur because of(1); occur during(1); occur at(1); be generated by(1); be found in(1); invade(1); occur from(1); involve(1); be related to(1)

Juvenile court [FOR] try(8); handle(3); deal with(2); prosecute(2); involve(2); process(2); adjudicate(1); adjudicates for(1); govern(1); relate to(1); cater to(1); punish(1); specialize in(1); pertain to(1)

Kennel puppies [FROM] live in(8); come from(5); be raised in(4); be born in(4); be found in(2); be bred in(2); be from(2); be kept in(1)

Lace handkerchief [HAVE₁] be made of(16); be made from(4); be trimmed with(1); be made out of(1); be comprised of(1); be trimmed in(1); come from(1); be decorated with(1); have(1)

Language riots [ABOUT] involve(5); be caused by(4); concern(2); be about(2); be incited because of(1); be caused for(1); be inspired by(1); occur because of(1); happen because of(1); stem from(1); pertain to(1); result from(1); be rooted in(1); be because of(1)

Laugh wrinkles [CAUSE₂] be caused by(12); come from(8); result from(2); show as(1); be gotten from(1); be created by(1); occur from(1); appear with(1)

Lemon peel [HAVE₂] come from(17); be found on(2); cover(2); come off(2); be on(1); be made from(1); protect(1); surround(1)

Lightning rod [FOR] attract(14); conduct(3); channel(3); direct(2); arrest(2); draw(1); safeguard from(1); deflect(1); contain(1); invite(1); protect from(1)

Lion cub [BE] come from(7); be(5); become(4); grow into(3); be birthed by(3); be born of(2); be sired by(1); bear of(1); be born from(1); descend from(1)

Love song [ABOUT] be about(12); speak of(3); talk about(2); be written about(1); describe(1); profess(1); speak about(1); express(1); sing about(1); concern(1); talk of(1); nourish(1); imply(1)

Machine translation [USE] be done by(10); be performed by(4); be made by(3); come from(3); be generated by(1); generate from(1); emanate from(1); pertain to(1); be developed by(1); be specific for(1)

Marine life [IN] be(7); exist in(2); be lived by(1); occur in(1); surprise(1); turn(1); refer to(1); emanate from(1); be led by(1); characterize(1); live in(1); be found in(1); involve(1); be related to(1)

Midnight snack [IN] be eaten at(15); occur at(3); be consumed at(2); happen at(2); be eaten around(1); occur around(1); around(1); be had at(1); eat in(1); go into(1); happen(1); be at(1)

Milieu therapy [USE] involve(3); use(2); describe(1); occur in(1); be based upon(1); change(1); consider(1); be based on(1); host(1); be performed in(1); be changed by(1); work with(1); employ(1); practice(1); be used on(1); be(1); happen in(1); require(1); concern(1)

Mining engineer [FOR] specialize in(4); study(3); know about(3); do(2); focus on(2); specialise in(1); understand(1); work with(1); know(1); work in(1); investigate(1); research(1); work at(1); relate to(1); work on(1); be involved in(1)

Morning prayers [IN] be said in(8); be done in(2); be said(2); occur at(2); be made in(2); happen in(2); be recited in(1); be chanted in(1); be recited during(1); be meant for(1); occur in(1); happen during(1); terminate in(1); be given during(1); be spoken in(1)

Morphology lecture [ABOUT] be about(9); describe(4); discuss(2); include(2); explain(2); deal with(1); address(1); teach(1); teach about(1); cover(1); concern(1)

Moth hole [CAUSE₂] be made by(12); be caused by(3); be eaten by(3); be created by(2); come from(2); protect(1); be produced by(1); be eaten out by(1)

Mother church [BE] act as(6); be(3); act like(3); resemble(2); be like(1); be originated with(1); originates from(1); recruit(1); remind of(1); operate as(1)

Mountain lodge [IN] be located in(7); be in(7); be situated on(2); be built on(1); be on(1); be built in(1); be located on(1); be situated in(1); be up in(1); sit in(1)

Mountain range [MAKE₂] be made of(5); contain(4); be composed of(2); consist of(2); be of(2); be(1); be created by(1); be comprised of(1); run through(1); be made up of(1)

Movement schisms [IN] divide(2); involve(2); split(1); occur to(1); emerge from(1); occur in(1); form within(1); be created by(1); be characterized by(1); separate(1); result in(1); emanate from(1); be due to(1); come from(1); have(1); be caused by(1); result from(1); be related to(1); be created from(1); affect(1); require(1); concern(1); within(1)

Murder charge [BE] involve(10); relate to(2); consist of(2); concern(2); be for(1); include(1); be labeled(1); indicate(1); come from(1); be of(1); claim(1); be related to(1); presuppose(1); be concerned with(1)

Music box [MAKE₁] play(15); produce(3); make(3); include(1); contain(1); be made for(1); deliver(1)

Neighborhood bars [IN] be in(8); be located in(4); be found in(2); be situated in(1); protect(1); exist within(1); cater to(1); encircle(1); reside near(1); reside in(1); be housed in(1); belong in(1)

Night flight [IN] occur at(10); happen at(7); go at(2); fly during(1); fly at(1); begin at(1); fly in(1); happen by(1); originate in(1); come at(1); travel at(1)

Nose drops [FOR] be used in(4); go in(4); come from(3); be placed in(2); treat(2); go into(2); be for(1); drain from(1); be inserted into(1); drop into(1); clear(1); relieve(1); help(1); be intended for(1); moisturize(1); heal(1)

Novelty item [BE] be(11); display(1); be seen as(1); be considered(1); be sold as(1); represent(1); become(1); be loved for(1); have(1); resemble(1); be regarded as(1); seem(1); consist of(1); generate(1)

Nut bread [HAVE₁] contain(11); be made with(2); consist of(2); be made of(2); be made using(2); include(1); be filled with(1); have(1); be composed of(1); be made from(1); come from(1)

Office friendships [IN] start in(2); form in(2); begin in(2); occur in(2); be in(2); be formed in(2); begin at(2); be made in(1); originate in(1); flourish in(1); pertain to(1); originate at(1); happen at(1); come from(1); reside in(1); occur at(1); be cultivated in(1); start at(1); happen in(1); start from(1)

Oil crisis [ABOUT] involve(7); be caused by(6); be about(2); depend upon(1); relate to(1); be precipitated by(1); deal with(1); be due to(1); be related to(1); be from(1); overuse(1); concerns(1); squeeze(1); affect(1)

Oil well [FOR] contain(9); produce(6); pump(3); draw(2); spew(1); have(1); dispense(1); pump up(1); harness(1)

Olive oil [FROM] come from(11); be made from(4); be made of(3); be derived from(1); be pressed from(1); be(1); smell of(1); contain(1); be extracted from(1); be pressed out of(1)

Onion tears [CAUSE₂] be caused by(14); come from(3); be generated by(1); be created by(1); arise due to(1); be induced by(1); be provoked by(1); react to(1); flow with(1)

Paper money [MAKE₂] be made of(11); be made from(8); be(3); be printed on(2); be composed of(1); be made out of(1); be produced on(1)

Parent organization [BE] be composed of(4); be made up of(4); be for(2); be run by(2); consist of(2); include(1); involve(1); be(1); supervise like(1); be made for(1); be characterized as(1); invite(1); oversee like(1); be filled with(1)

Party members [HAVE₂] belong to(10); form(2); attend(2); go to(2); join(1); be in(1); participate in(1); come from(1); be of(1); make up(1); be involved with(1); belong with(1); interact at(1)

Peanut butter [FROM] be made from(10); be made of(5); come from(5); contain(3); be processed from(2); be made with(1); be composed of(1); be used with(1); be formulated from(1)

Pedal extremities [BE] be(6); be used to(2); resemble(2); use(2); occur on(1); move(1); operate(1); include(1); be located on(1); look like(1); contain(1); maneuver(1); move like(1); be from(1); require(1)

People power [HAVE₂] come from(7); belong to(3); emerge from(1); reside with(1); be made by(1); be within(1); be exercised by(1); be used by(1); reside in(1); drive(1); involve(1); result from(1); motivate(1); be given to(1); be due to(1); be shown by(1); encompass(1)

Pet families [HAVE₁] have(9); own(5); be made of(2); like(1); care for(1); keep(1); enjoy(1); consist of(1); adopt(1); love(1); be(1); be composed of(1); be made up of(1); be formed from(1)

Pet spray [FOR] be used on(7); be for(2); go on(2); deter(1); be exuded by(1); be on(1); be made for(1); go onto(1); relate to(1); be applied to(1); come from(1); cover(1); consist of(1); touch(1); aid(1); deodorize(1); be applied on(1); be designed for(1); deodorizes(1)

Pet theory [BE] involve(5); be about(3); pertain to(2); be(1); apply to(1); be treated like(1); refer

to(1); relate to(1); resemble(1); be favored like(1); be like(1); be regarding(1); be considered(1); concern(1)

Phantom limb [BE] be(7); act like(3); be like(2); feel like(2); have(1); appear(1); behave like(1); mimic(1); look like(1); exist as(1); resemble(1)

Picture album [FOR] contain(18); hold(7); include(1); be filled with(1); have(1)

Picture book [HAVE₁] contain(15); have(3); be composed of(1); rely on(1); use(1); consist of(1); include(1); contains(1); be made of(1)

Pine tree [BE] have(4); smell like(3); be made of(3); be(3); be called(2); smell of(2); produce(2); be made from(1); belong to(1); be named(1); classify as(1); relate to(1); bear(1); contain(1)

Plant food [FOR] nourish(7); feed(7); be for(5); be made for(1); be used by(1); benefit(1); be given to(1); be given(1); be absorbed by(1); be supplied by(1); be eaten by(1); nurtures(1); be fed to(1)

Plastic toys [MAKE₂] be made of(15); be made from(4); contain(2); consist of(1); be(1); be formulated of(1); be manufactured from(1)

Plum wine [FROM] be made from(15); come from(5); be made of(2); be made with(2); contain(1); taste like(1); be produced from(1)

Pole height [HAVE₂] measure(3); reach(2); be set for(1); be reached by(1); be ascribed to(1); include(1); limit(1); come from(1); be measured on(1); determine(1); occur within(1); be like(1); be compared to(1); be measured for(1); reach to(1); involve(1); be related to(1); of(1)

Policy matters [ABOUT] involve(4); concern(4); relate to(2); pertain to(2); be related to(2); deal with(1); be of(1); have(1); regard(1); affect(1); be relative to(1); belie to(1); refer to(1); be created as(1); use(1); reference(1); follow(1); be answered by(1); be about(1)

Pork suet [FROM] be made from(10); be made of(5); come from(4); be rendered from(2); be trimmed from(1); use(1); contain(1); be from(1); make of(1)

Pot high [CAUSE₂] come from(9); be caused by(3); result from(3); be derived from(1); utilize(1); be created by(1); derive from(1); be induced by(1); be from(1); be created from(1); be produced from(1)

Pressure cooker [USE] use(9); cook with(6); utilize(2); cook under(2); produce(1); generate(1); build up(1); work with(1); operate use(1); involve(1); cook by(1); cook use(1)

Price dispute [ABOUT] involve(8); be over(5); concern(3); be about(3); arise over(1); argue about(1); center on(1); question(1); argue over(1); arise from(1)

Queen bee [BE] be(13); act as(2); grow into(1); look like(1); act like(1); rule as(1); come from(1)

Radio communication [USE] come from(4); use(3); be transmitted over(1); be broadcast by(1); be done over(1); flow through(1); emerge from(1); transmit to(1); be made by(1); be by(1); emanate from(1); be transmitted by(1); happen over(1); consist of(1); involve use(1); occur via(1); be made via(1); be made with(1); come through(1); be transmitted through(1)

Rice paper [FROM] be made of(12); be made from(4); make from(1); contain(1); be produced from(1); be made out of(1); be composed of(1); in make from(1); be made up of(1); make of(1); come from(1); consist of(1); be from(1)

Rye whiskey [FROM] be made from(14); be made of(4); come from(3); be made with(2); be distilled from(1); be composed of(1); be made using(1)

Salt lake [HAVE₁] contain(14); have(6); be made of(2); be made from(2); include(1); be composed

of(1); be high in(1)

Sand dune [MAKE₂] be made of(11); be formed of(2); contain(2); be composed of(1); be formed by(1); be formed from(1); consist of(1); originate from(1); be made from(1); be comprised of(1); be made up of(1); build from(1); be from(1); be composed from(1); be build by(1); be filled with(1)

Sap tree [MAKE₁] produce(12); have(4); supply(2); give(2); exude(1); make(1); grow(1); create(1); ooze(1); ooze out(1)

Satellite nation [BE] be(4); act like(3); consist of(2); have(1); be like(1); use(1); depend on like(1); be composed of(1); be made up of(1); look like(1); operate like(1); serve as(1); revolve like(1); resemble(1); rely on(1); supply(1); receive from(1)

Sea breeze [FROM] come from(21); blow off(1); come off(1); be near(1); come off of(1)

Servant girl [BE] be(10); work as(8); act as(5); be used as(1); be employed as(1); become(1); exemplify(1); come from(1)

Sex scandal [ABOUT] involve(12); be about(4); be caused by(2); have(1); center around(1); relate to(1); deal with(1); be made because of(1); concern(1); be concerned with(1); require(1)

Shock treatment [USE] use(10); involve(2); cause(2); include(1); induce(1); provide(1); involves(1); be comprised of(1); administer(1); contain(1); be based on(1); utilize(1); consist of(1); require(1)

Sister node [BE] be(3); accompany(2); act as(2); have(1); refer to(1); be related like(1); pertain to(1); be of(1); work with(1); be like(1); make(1); act like(1); be related to(1); be likened to(1); behave as(1); duplicate(1)

Smoke signals [USE] be made of(9); be made from(5); use(3); emerge from(2); be(1); be caused by(1); consist of(1); be comprised of(1); result in(1); signal(1); come from(1); resemble(1)

Snow blindness [CAUSE₂] be caused by(22); come from(3); be created by(1); exemplify(1)

Sob story [CAUSE₁] make(5); cause(5); evoke(4); induce(3); elicit(3); produce(2); make for(1); provoke(1); cause to(1); be made of(1); illicit(1)

Soldier ant [BE] act like(10); be(4); behave like(4); act as(2); work like(2); march on like(1); mimic(1); emulate(1); play(1); serve as(1); resemble(1)

Song bird [MAKE₁] sing(19); make(2); produce(1); warble(1); love(1); erupt in(1)

Sports activities [BE] involve(8); include(3); be comprised of(2); be(2); revolve around(1); be composed of(1); center around(1); come from(1); depend on(1); work on(1); be found in(1); be related to(1); concern(1); be concerned with(1)

Sports magazine [ABOUT] feature(5); discuss(3); be about(3); cover(2); focus on(2); tell about(2); deal with(1); involve(1); write on(1); review(1); refer to(1); contain(1); be found in(1); detail(1); be interested in(1); highlight(1)

Star shape [BE] resemble(12); look like(10); be like(1); recall(1)

Starvation diet [USE] cause(3); mimic(2); require(2); lead to(2); come from(2); be like(2); induce(2); use(2); involve(1); be(1); feel like(1); include(1); generate from(1); necessitate(1); simulate(1); allow(1); work by(1); help in(1)

Steam iron [USE] use(7); produce(6); emit(3); utilize(3); make(1); create(1); incorporate(1); steam(1); puff(1); vent(1); give off(1); release(1)

Steel helmet [MAKE₂] be made of(15); be made from(6); be created of(1); be made out of(1);

contain(1); come from(1)

Stone tools [MAKE₂] be made of(15); be made from(5); work with(1); shape(1); cut(1); carve away(1); work on(1)

Store clothes [FROM] be bought in(6); come from(4); be sold in(2); be found in(2); be purchased in(2); be bought from(2); be used in(1); be from(1); be located in(1); be bought(1); be purchased from(1); buy(1); be bought at(1)

Student committee [MAKE₂] be comprised of(5); be made up of(5); consist of(3); contain(3); be made of(3); be composed of(2); involve(2); include(1); represent(1); be headed by(1)

Student friends [BE] be(15); be had by(2); meet as(1); help(1); be considered(1); come from(1); be made by(1); exist as(1); comprise(1); matriculate as(1)

Student power [HAVE₂] come from(5); be possessed by(3); be held by(2); involve(2); be exercised by(2); inspire(1); be wielded by(1); belong to(1); be yielded by(1); derive from(1); emanate(1); be expressed by(1); relate to(1); devolve from(1); be seized by(1); be of(1); be exerted by(1); be about(1)

Student problems [HAVE₂] be had by(4); concern(3); involve(2); affect(2); belong to(2); occur with(1); originate from(1); be encountered by(1); frustrate(1); plague(1); trouble(1); come from(1); be caused by(1); occur(1); worry(1); face challenge(1); happen to(1); be possessed by(1); be found in(1); burden(1); be owned by(1)

Sugar cube [MAKE₂] be made of(14); be made from(5); consist of(3); be formed from(2); contain(1); be composed of(1); be comprised of(1); be all(1); store(1); be formed of(1)

Summer months [IN] be in(7); occur in(5); occur during(3); happen in(2); begin in(1); appear in(1); come in(1); be during(1); contain(1); be included in(1); fall in(1); comprise(1); happen during(1)

Summer travels [IN] occur during(10); happen in(5); occur in(3); be in(2); enhance(1); occur(1); be made in(1); involve(1); begin during(1); go into(1); happen during(1); be done in(1)

Surface tension [IN] be on(5); occur on(3); occur at(2); reside on(2); appear on(1); disrupt(1); exist on(1); affect(1); be found on(1); be experienced on(1); be formed on(1); influence(1); retain(1); shape(1); be measured at(1); be felt on(1); cover(1); form at(1); lie across(1); hold across(1)

Suspense film [CAUSE₁] contain(7); have(5); create(4); be(2); cause(2); build(1); include(1); provide(1); generate(1); be filled with(1); give(1); s purpose be(1); supply(1); be about(1)

Tape measure [BE] use(4); be made of(4); be made from(2); be on(1); be in(1); be made by(1); be created out of(1); look like(1); be formed by(1); contain(1); run along(1); resemble(1); be like(1); be provided by(1); consist of(1); be shaped like(1); be written on(1)

Target structure [BE] be(10); have(3); include(2); hold(1); represent(1); contain(1); become(1); aim at(1); resemble(1); uphold(1); be aimed at(1); be shot at like(1)

Tax law [ABOUT] be about(6); involve(5); govern(2); deal with(2); pertain to(2); be related to(2); justify(1); define(1); refer to(1); relate to(1); dictate(1); focus on(1); explain(1)

Teaching profession [BE] involve(12); require(3); include(2); espouse(2); perform(1); enjoy(1); specialize in(1); revolve around(1); be devoted to(1); be about(1); focus upon(1)

Tear gas [CAUSE₁] cause(15); create(2); produce(1); generate(1); shades(1)

Testtube baby [FROM] be conceived in(12); be made in(4); be created in(4); come from(4); be born

from(1); be fertilized in(1); be created with(1)

Tire rim [HAVE₂] surround(4); hold(3); be on(3); fit(2); cover(2); be of(2); protect(1); be in(1); include(1); involve(1); accept(1); be surrounded by(1); lock in(1); be included with(1); be within(1); be located in(1); circle(1); be inside of(1)

Tobacco ash [FROM] come from(15); emerge from(2); derive from(2); stick to(1); be made from(1); result from(1); be(1); be from(1); be created from(1); be produced from(1)

Vacuum cleaner [USE] use(5); clean by(2); utilize(2); be(1); do(1); suck into(1); suck by(1); pick up with(1); work on(1); work with(1); employ(1); function as(1); need(1); work via(1); suck in(1)

Vapor lock [CAUSE₂] be caused by(7); contain(4); result from(3); control(2); be created by(1); form(1); seal in(1); come from(1); develop in(1)

Vegetable soup [HAVE₁] contain(8); be made from(5); be made of(3); come from(3); consist of(1); have(1); be composed of(1); have within(1); be comprised of(1); be made up of(1); be made using(1)

Voice vote [USE] be made by(7); be given by(2); use(2); be done by(2); be taken by(1); be conducted by(1); count(1); occur through(1); determine by(1); be expressed by(1); be cast using(1); come from(1); be counted by(1); rely on(1); need(1); require(1)

Wastebasket category [BE] belong in(4); include(3); go in(2); describe(1); be named(1); deal with(1); contain(1); relegate to(1); be named after(1); use(1); provide(1); involve(1); emulate(1); be related to(1); trash into(1); go into(1); be about(1)

Water drop [MAKE₂] be made of(7); be composed of(3); consist of(3); be comprised of(2); be(1); contain(1); have(1); emerge from(1); swim in(1); hold(1); be formed from(1); come from(1)

Water mark [CAUSE₂] be made by(5); be caused by(3); indicate(2); look like(2); be made from(2); be created by(2); show(1); emboss(1); delineate(1); stain(1); be left by(1); come from(1); be made of(1); stain like(1); be made with(1)

Water wheel [USE] move(4); use(3); turn use(2); spin(1); turn by(1); turn with(1); utilize(1); work in(1); be composed of(1); run on(1); tumble(1); turn from(1); flow with(1); transport(1); reside in(1); work with(1); take out(1)

Weekend boredom [IN] occur on(10); happen during(4); occur during(3); happen on(3); emerge at(1); be produced by(1); mar(1); characterize(1); experience in(1); come on(1)

Winter season [BE] include(4); be(4); occur in(2); be in(2); occur during(2); come in(2); describe(2); consist of(2); coincide with(1); feel like(1); be called(1); cover(1); exemplify(1); contain(1); recall(1); occur at(1); comprise(1)

Winter sports [IN] be played in(11); occur in(5); happen in(3); be played during(2); happen during(2); be done in(1); utilize(1); be in(1); be enjoyed in(1); occur during(1)

Women professors [BE] be(21); teach(1); look like(1); resemble(1)

Wood shavings [FROM] come from(7); be made of(7); be made from(4); contain(1); emerge from(1); be created by(1); be comprised of(1); be formed from(1); be formed by(1); be carved off(1); be carved from(1); come out of(1); be produced from(1)

Worker teams [MAKE₂] be comprised of(5); be made up of(4); be composed of(2); include(2); have(2); comprise(2); contain(1); use(1); bring(1); consist of(1); involve(1); incorporate(1); be made of(1); be comprised by(1)

Tableau D.1 – suite

i	Nom	Élément extrait	Expression régulière
31	clause relative, gr. prépositionnel, particule, tête seulement	verbe + particule	<pre>(?:[^\:]*:[^\:]*NP:word:[^_]*_?(?:IN:[^\:]*PP:[^\:]*:[^_]*_ [^\:]*:[^\:]*NP:[^\:]*:[^_]*_?(?:[^\:]*:[^\:]*NP:[^\:]*: [^_]*_)**(?:(:WDT) (?:WP))(:[^\:]*:[^\:]*:[^_]*_)(?: [^\:]*:[^\:]*VP:[^\:]*:[^_]*_)*(VB[^\:]*:[^\:]*VP:[^\:]*: [^_]*_)((?:(:RP) (?:IN)): [^\:]*:[^\:]*:[^_]*_)(?:[^\:]*: [^\:]*NP:[^\:]*:[^_]*_)*(?:[^\:]*:[^\:]*NP:word:[^_]*)(?: _[^\:]*:(?:(!I-NP)[^\:]*\.\.)*)*\$</pre>
32	clause relative, gr. prépositionnel, particule, tête seulement, passif	be + verbe + particule	<pre>(?:[^\:]*:[^\:]*NP:word:[^_]*_?(?:IN:[^\:]*PP:[^\:]*: [^_]*_[^\:]*:[^\:]*NP:[^\:]*:[^_]*_?(?:[^\:]*:[^\:]*NP:[^\:]*: [^_]*_)**(?:(:WDT) (?:WP))(:[^\:]*:[^\:]*:[^_]*_)(?: [^\:]*:[^\:]*VP:[^\:]*:[^_]*_)*([^\:]*:[^\:]*be:[^_]*_ (VB[^\:]*:[^\:]*VP:[^\:]*:[^_]*_)((?:(:RP) (?:IN)): [^\:]*: [^\:]*:[^_]*_)(?:[^\:]*:[^\:]*NP:[^\:]*:[^_]*_)*(?:[^\:]*: [^\:]*NP:word:[^_]*)(?:_[^\:]*:(?:(!I-NP)[^\:]*\.\.)*)*\$</pre>

ANNEXE E Noms composés de Kim classés selon les 12 prédicats de Levi

- | | |
|--|--|
| 1. accessories portion : <i>for, about</i> | 33. bathing suit : <i>for</i> |
| 2. accounting controversy : <i>about</i> | 34. bed lump : <i>in</i> |
| 3. acquisition agreement : <i>about</i> | 35. bedrock sweeper : <i>in</i> |
| 4. ad clutter : <i>have₁</i> | 36. benchmark issue : <i>have₂</i> |
| 5. advance look : <i>have₁, about</i> | 37. bidding war : <i>in</i> |
| 6. advertising agency : <i>for</i> | 38. blood test : <i>use</i> |
| 7. advertising blitz : <i>for</i> | 39. board member : <i>have₂</i> |
| 8. advertising firm : <i>for</i> | 40. bond investor : <i>use, have₁</i> |
| 9. agriculture trade : <i>about, from</i> | 41. bond price : <i>have₂</i> |
| 10. aid package : <i>for</i> | 42. bottom line : <i>in</i> |
| 11. air carrier : <i>in</i> | 43. brand image : <i>have₂</i> |
| 12. aircraft series : <i>have₁</i> | 44. brat pack : <i>have₁</i> |
| 13. airline company : <i>make₁</i> | 45. brokerage network : <i>about, for</i> |
| 14. aluminum unit : <i>make₂</i> | 46. brokerage service : <i>for</i> |
| 15. anthrax death : <i>cause₂</i> | 47. budget gimmick : <i>about</i> |
| 16. anthrax outbreak : <i>cause₂</i> | 48. budget impact : <i>cause₂, have₂</i> |
| 17. apartment complex : <i>have₁</i> | 49. budget pressure : <i>cause₂, from</i> |
| 18. apparel industry : <i>make₁</i> | 50. budget uncertainty : <i>about</i> |
| 19. applause sign : <i>for, about</i> | 51. bull trend : <i>be</i> |
| 20. area hospital : <i>in</i> | 52. business activity : <i>be</i> |
| 21. assembly plant : <i>for</i> | 53. business category : <i>about</i> |
| 22. assembly worker : <i>for</i> | 54. business executive : <i>in</i> |
| 23. auto demand : <i>have₂</i> | 55. business loan : <i>for</i> |
| 24. automobile part : <i>make₁, from</i> | 56. business school : <i>for, about</i> |
| 25. backup system : <i>for</i> | 57. business sector : <i>make₂</i> |
| 26. balance due : <i>be</i> | 58. buying interest : <i>for</i> |
| 27. bank credit : <i>from</i> | 59. buying opportunity : <i>for</i> |
| 28. bank financing : <i>from</i> | 60. buying survey : <i>about</i> |
| 29. bank lender : <i>be</i> | 61. cable partner : <i>have₁</i> |
| 30. banking company : <i>about, make₁</i> | 62. cable service : <i>use</i> |
| 31. bankruptcy court : <i>for, about</i> | 63. cable system : <i>make₂</i> |
| 32. bankruptcy proceeding : <i>for</i> | 64. capita income : <i>have₂</i> |
| | 65. capture rate : <i>have₂</i> |
| | 66. car dealership : <i>about, use</i> |

67. car factory : *for*
68. cash grants : *be*
69. cash reserve : *make₂*
70. cast actor : *in*
71. center stage : *in*
72. cereal maker : *make₁*
73. chemical industry : *make₁*
74. circulation guarantee : *for*
75. circulation level : *have₂*
76. clock radio : *have₁, be*
77. cocktail party : *have₁*
78. collection agency : *make₁*
79. companion jetty : *be*
80. company lawyer : *have₂*
81. company plan : *for*
82. compensation insurance : *for*
83. computer printer : *for*
84. computer retailer : *for*
85. consensus forecast : *be*
86. consent decree : *about, be*
87. consent process : *for*
88. consulting giant : *for*
89. consumer demand : *make₂, from*
90. consumer magazine : *for*
91. consumer market : *about, have₁*
92. consumer product : *for*
93. consumer research : *about*
94. contract terms : *in, have₂*
95. control system : *for*
96. convenience phone : *be*
97. cookie baker : *make₁*
98. cooperation agreement : *about, for*
99. core rate : *about, for*
100. country house : *in*
101. crop price : *have₂*
102. crude production : *make₁*
103. cruise ship : *for*
104. crystal chandelier : *make₂*
105. currency dealing : *use*
106. customer base : *be*
107. customer inventory : *in, have₂*
108. customer repurchase : *make₂*
109. customer satisfaction : *have₂*
110. dashboard computer : *in*
111. data storage : *for*
112. data terminal : *for*
113. defense contractor : *for*
114. defense designate : *for*
115. defense program : *for*
116. department spokeswoman : *from*
117. department store : *have₁*
118. deposit note : *for*
119. depository institution : *use, have₁*
120. design work : *be, make₁*
121. development company : *for*
122. dieting advice : *about*
123. dinner discussion : *in, be*
124. disaster area : *have₂*
125. disaster loan : *for*
126. disk drive : *be, in*
127. distribution facility : *for*
128. dividend yield : *make₂, from*
129. dream world : *be*
130. drilling rig : *make₁, for*
131. drug manufacturer : *make₁*
132. earning prospect : *about*
133. earthquake loss : *cause₂*
134. earthquake relief : *for*
135. economics journalist : *about*
136. education service : *about, cause₁*
137. election district : *for*
138. electronics problem : *in, cause₂*
139. emergency funding : *for*

140. employee manual : *have₂*
141. employment act : *about*
142. end product : *in, be*
143. engineering company : *make₁*
144. entertainment company : *for*
145. entertainment series : *have₁*
146. entertainment show : *be, for*
147. entitlement program : *for*
148. equity market : *for*
149. equity stake : *for*
150. estate concern : *about*
151. estate magnate : *about*
152. exchange commission : *about, for*
153. exchange market : *for*
154. exploration side : *about*
155. extrusion division : *make₁, for*
156. face value : *from, have₂*
157. factory consolidation : *in, have₂*
158. factory owner : *have₁*
159. family album : *about, have₂*
160. fashion magazine : *about*
161. federal fund : *from*
162. film festival : *about*
163. finance opportunity : *for*
164. fire insurance : *about*
165. flavor packet : *have₁*
166. forest product : *from*
167. freeway survivor : *from*
168. front yard : *in*
169. fuel cost : *have₂*
170. future deal : *in*
171. future demand : *in*
172. future gain : *in*
173. future loss : *in*
174. future merger : *in*
175. gallon tax : *cause₂, for*
176. game fishing : *be*
177. gas division : *for, about*
178. gas producer : *make₁*
179. gas well : *have₁*
180. giveaway program : *make₁*
181. glass bottle : *be, make₂*
182. government hand : *have₂*
183. government loss : *have₂*
184. government official : *for, from*
185. government plan : *from*
186. government research : *make₂*
187. governments authority : *be*
188. graduate student : *be*
189. graphics computer : *have₁, for*
190. group plan : *for*
191. highway assistance : *in*
192. horse barn : *for, have₁*
193. horse breeder : *for*
194. hospital room : *in*
195. hotel comedy : *in*
196. hype claim : *cause₁*
197. impeachment trial : *about, for*
198. industry executive : *have₂, for*
199. industry relation : *have₂*
200. industry standard : *have₂*
201. inflation threat : *be, cause₂*
202. information service : *about, for*
203. insurance analyst : *for*
204. insurance commissioner : *for*
205. insurance issue : *about*
206. interest expense : *cause₂, be*
207. interest rate : *have₂*
208. investment bank : *for, make₁*
209. investment banker : *make₁*
210. investment banking : *for*
211. investment vehicle : *for*
212. iron stake : *make₂*

213. issues forecast : *about*
214. ivory tower : *make₂*
215. jet engine : *in, have₂, for*
216. junk bond : *be*
217. junk issue : *be*
218. keyboard instrument : *have₁, use*
219. lawn party : *in*
220. lead character : *be, in*
221. leisure market : *about*
222. lending program : *about, make₁*
223. liquor industry : *make₁*
224. listing requirement : *for, have₂*
225. loan origination : *have₂, for*
226. loan program : *about*
227. loan representative : *for*
228. machining center : *make₁*
229. maintenance expense : *for, cause₂*
230. management differences : *in, about*
231. management lapse : *have₂*
232. manufacture executive : *in, for*
233. mark bond : *use, be*
234. market activity : *make₂*
235. market gyration : *make₂*
236. market research : *about*
237. market segmentation : *have₂*
238. marketing sector : *about*
239. material supply : *be*
240. media director : *have₂*
241. member remark : *make₂*
242. merchandise store : *for*
243. metals analyst : *for, about*
244. metals sector : *about*
245. milligram version : *in, be*
246. minority stake : *be*
247. minute description : *have₁, about*
248. missile contract : *about*
249. morning tea : *in*
250. motion picture : *have₁*
251. motor vehicle : *use*
252. murder mystery : *about*
253. mystery novel : *about, have₁*
254. narcotics trafficker : *about*
255. negotiator leverage : *have₂*
256. network television : *be*
257. news magazine : *about*
258. nonfiction television : *about*
259. office system : *have₂, for*
260. oil company : *make₁, have₁*
261. oil exploration : *for*
262. oil spill : *make₂, be*
263. operation officer : *have₂*
264. operation profit : *have₂*
265. parent company : *be*
266. parent corporation : *be*
267. passenger car : *for*
268. patent case : *about*
269. patent law : *about*
270. payment deadline : *for*
271. payroll tax : *in, about*
272. percent difference : *in, about, be*
273. percentage point : *be*
274. percentage rate : *be*
275. petrochemical corridor : *for*
276. phone company : *about*
277. planning discipline : *about, in*
278. planning group : *make₁*
279. plant delay : *cause₂, in*
280. plant equipment : *for*
281. plumbing authority : *about*
282. policy squabble : *about*
283. pollution control : *for*
284. pool hall : *have₁*
285. post times : *in*

- 286. power source : *make*₁
- 287. press coverage : *from*
- 288. price action : *make*₂
- 289. price data : *about*
- 290. price deterioration : *have*₂
- 291. price increase : *have*₂
- 292. price quote : *about*
- 293. price range : *from*
- 294. price rise : *have*₂
- 295. price swing : *make*₂
- 296. price tag : *for, about*
- 297. pricing policy : *about*
- 298. pricing program : *for*
- 299. pricing strategy : *for*
- 300. problem asset : *be, cause*₁
- 301. processing service : *make*₁, *be*
- 302. production earning : *have*₂
- 303. production profit : *have*₂
- 304. profit growth : *have*₂
- 305. profit rise : *have*₂
- 306. property claim : *about*
- 307. property portfolio : *have*₁, *for*
- 308. property rate : *have*₂
- 309. quarter earning : *in*
- 310. racing program : *about*
- 311. racketeering prosecution : *about*
- 312. racketeering suit : *about*
- 313. rail depot : *have*₁
- 314. ramp worker : *in*
- 315. range version : *have*₁
- 316. rate reduction : *have*₂
- 317. recovery cost : *have*₂, *for*
- 318. redemption price : *for, about*
- 319. reelection campaign : *for*
- 320. reorganization plan : *for*
- 321. reporting period : *for*
- 322. research department : *about*
- 323. research division : *make*₁
- 324. research scientist : *make*₁
- 325. resin derivative : *from*
- 326. rights group : *about*
- 327. risk assessment : *for, about*
- 328. robot spacecraft : *be*
- 329. safety standard : *for*
- 330. sales rate : *have*₂
- 331. sales report : *about*
- 332. school chief : *have*₂, *for*
- 333. sea wall : *for*
- 334. search firm : *make*₁
- 335. section volume : *have*₂
- 336. seesaw gyration : *have*₂
- 337. selling pressure : *have*₂, *about*
- 338. service company : *for*
- 339. service effort : *for*
- 340. severance cost : *cause*₂, *make*₂
- 341. sex symbol : *about*
- 342. share point : *about*
- 343. shareholder value : *have*₂
- 344. shelf space : *in*
- 345. shipyard building : *be*
- 346. soil fungus : *in*
- 347. spending bill : *about*
- 348. spending cash : *for*
- 349. spending figure : *about*
- 350. spin-off series : *be*
- 351. spirits division : *about, have*₁
- 352. state bar : *from*
- 353. state delegation : *have*₂
- 354. state official : *have*₂
- 355. state senate : *for*
- 356. steel market : *for, about*
- 357. steel trade : *about, from*
- 358. stock character : *from*

359. stock fluctuation : *make₂*
 360. stock market : *for*
 361. stock slump : *have₂*
 362. storage facility : *for*
 363. store security : *for*
 364. story consultant : *for*
 365. stretch bra : *make₂, have₁*
 366. stub stock : *be*
 367. studio audience : *in*
 368. subcommittee chairman : *have₂*
 369. subsidiary partnership : *be*
 370. summer sale : *in*
 371. takeover bid : *be*
 372. takeover offer : *be*
 373. takeover plan : *for*
 374. talk show : *use, have₁*
 375. tax credit : *for*
 376. tax rate : *have₁*
 377. teaching company : *for*
 378. technology issue : *cause₂, about*
 379. television ad : *use*
 380. television network : *have₁, use*
 381. television production : *for*
 382. thrift executive : *for*
 383. tire maker : *make₁*
 384. tissue maker : *make₁*
 385. titanium engine : *make₂*
 386. trade figure : *about*
 387. trade gap : *have₁*
 388. trade issue : *about, cause₂*
 389. trade monopoly : *about, in*
 390. trade number : *for, have₂*
 391. trading loss : *cause₂, from*
 392. trading partner : *for*
 393. trading strategy : *for, about*
 394. treasury bond : *make₂, from*
 395. truck market : *about*
 396. turnaround business : *make₁*
 397. turnaround expert : *cause₁, make₁*
 398. unemployment rate : *have₂, about*
 399. union election : *have₂, in*
 400. utility firm : *about, for*
 401. vanguard role : *have₂*
 402. vehicle rate : *have₁, for*
 403. vehicle sale : *have₁*
 404. veteran agent : *be*
 405. volume increase : *have₂*
 406. volunteer group : *be*
 407. voting power : *for*
 408. wage negotiation : *about*
 409. war hero : *in*
 410. warfare equipment : *for*
 411. watchdog agency : *be*
 412. weapons facility : *have₁*
 413. weekend purchase : *in*
 414. winnowing process : *be, cause₁*
 415. world report : *about*
 416. world series : *in*
 417. yield market : *have₁*